Naval Research Laboratory

Stennis Space Center, MS 39529-5004

Covariance Function for Nearshore Wave Assimilation Systems

JAY VEERAMONY

Ocean Dynamics and Prediction Branch
Oceanography Division

STYLIANOS FLAMPOURIS

University of Southern Mississippi
Stennis Space Center, Mississippi

MARK ORZECH
HANS NGODOCK

Ocean Dynamics and Prediction Branch
Oceanography Division

January 30, 2018

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-01-2018	Memorandum Report	

4. TITLE AND SUBTITLE

Covariance Function for Nearshore Wave Assimilation Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0602435N

6. AUTHOR(S)

Jay Veeramony, Stylianos Flampouris,[1] Mark Orzech, and Hans Ngodock

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
73-9170-V7-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/7320--18-9753

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

[1]Currently at NOAA/National Weather Service, National Center for Environmental Predictions, Environmental Modeling Center, College Park, Maryland.

14. ABSTRACT

Data assimilation is widely used to combine observations with dynamic models to improve model prediction. The relative weighting of the original forecast and the observations, the error covariance matrices, determine how the information is transferred from the observations to the model. It is shown that the optimization of the assimilation systems has to be driven by the data taking into account the physics of the wave field. Whereas the temporal covariance can be modeled by a parameterized Gaussian function, for nearshore wave assimilation applications, the covariance function depends primarily on the local depth and secondly on the distance from the assimilation location. In spectral space, a skewed exponential function is suggested.

15. SUBJECT TERMS

Data assimilation	Nearshore waves
Error covariance	

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	Unclassified Unlimited	64	Jay Veeramony
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited			**19b. TELEPHONE NUMBER** *(include area code)* (228) 688-4835

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

In earth sciences, data assimilation is a widely used technique to combine information from observations with dynamic models to improve model prediction. The technical details of the common methodologies are well documented, see e.g. Nichols (2010) for meteorological applications or Wunsch (2006) for oceanographic applications In short, the assimilation is based on the iterative minimization of the difference (called "*innovation*") between the forecast of a quantity (referred to as "*background*") at given location (of the appropriate space) and the measurement of the same quantity to produce a new, more accurate, model state (called "*analysis*"). In most assimilation schemes, the relative weighting of the background forecast and the observations is determined by a statistical estimate of the uncertainties in the model and observation data (Thornhill et al., 2012). The relative weightings, which is essentially the error covariance matrices, determine how the information is spread from the observations to the rest of the computational grid of the assimilation system and allow the two-way information "interaction" between the spot measurements and the simulated field. Regarding the covariance, Bennet claims that "[…] *it is difficult to develop covariances. It follows that the resulting inverse estimate or analysis of the circulation also lack credibility.*" (Bennett, 2002). Wunsch phrased his similar conclusion as: "[…] *the choice of the weights* (aka the covariance matrices) *dominates the current effort* (for assimilation). *At this stage, we have a problem primarily of oceanography and meteorology rather than one of mathematics.*"(Wunsch, 2006). This paper addresses the problem of generating the covariance matrix for wave data assimilation by focusing on the covariance of the background error of the wave energy. We present a general approach, driven by analysis of measurements and simulations with the SWAN wave model (Booij et al., 1999; Ris et al., 1999) which is applicable for any spectral wave model.

The four dimensional variational (4DVar) assimilation methods are based on the mathematical expression of the misfit between the model state and observations (known as cost function, J). It is typically expressed as

$$J(x) = \left(x_0 - x_0^b\right)^T B^{-1}\left(x_0 - x_0^b\right) + \sum_{i=1}^{K}(\psi_i - H(x_i^b))^T R_i^{-1}(\psi_i - H(x_i^b)) + \eta^T Q^{-1}\eta \quad (1.1)$$

where x_0 is the (wave) state at given time and location, x^b is the background state, ψ_i is the observation data, H corresponds to the observational operator (known also as measurement functional), B, R, Q are the background, observation and dynamic error covariances respectively. The basic mathematics for developing assimilation systems is known and fairly straightforward, but technically challenging (Bennett, 2002; Menke, 2012). However, determination of the covariances is an open question with multiple scientific challenges due to the number of possible errors from all the sources (instrument error and error of representativeness, external forcing errors, initial and boundary condition errors, errors in model parameterizations, unresolved processes and more) contained in them. The covariances are unknown and not typically measurable errors, and have to be estimated via assumptions, the adequacies of which are difficult to validate.

Data assimilation in wave models has been introduced relatively recently and mainly for large scale/ low resolution simulations, WAM, e.g. (Komen et al., 1994; Sannasiraj et al., 2006) and WAVEWATCH III (Chen et al., 2004; Tolman, 2009) which assimilate integrated spectral quantities, mostly only the significant wave height (H_s), mainly from satellite data. In addition to H_s, Aouf et al., (2006) assimilated into WAM other mean wave parameters such as direction and period. Feng et al., (2012) assimilated wave spectra from buoys into SWAN for the re-analysis of the wave field during typhoons in the East Pacific ocean. The assimilation significantly increased the accuracy of the significant wave height output, but the main disadvantage of their approach, judging by the output spectra, is that the directional and the frequency spread of the background (and not of the measured spectra) dominate the result.

The present investigation is the continuation of the first assimilation scheme for the nearshore wave model, SWAN, presented by Walker, (2006) for SAR data and applied with in situ data from Field Research Facility at Duck, validated by Veeramony et al., (2010), and it has several applications (Almeida et al., 2014). The data assimilation system by Walker is a strong constraint approach with the assumption that the model dynamics are perfect and the error sources are the imposed boundary conditions. In order to invert the spectral action equation and build the adjoint function, their method took into account only the wave propagation and neglected the source and sink terms of SWAN, which describe the non-linear phenomena of wave generation and decay. Hence, the problem was simplified to the inverse of a linear system, and an analytical adjoint function was determined. This approach is reasonable for deep ocean applications and propagation of swell, but it is insufficient for nearshore applications where the non-linear physics (wave breaking, bottom friction, etc.) are dominant. In addition, as it is discussed by Orzech et al., (2013), the discretization of the analytical model leads to erroneous numerical gradient of the cost function. Therefore Orzech produced a "numerical adjoint" of SWAN by inverting the discretized forward model and the "SWAN representer" function by linearizing the forward model in order to build a weak constraint 4D-Var assimilation system. As the assimilation system is based on the representer method (Bennett, 2002), it requires use of an appropriate covariance, e.g. Zupanski (1997), as weighting factors in spectral, spatial and temporal domain. The system is named SWAN-FAR.

The determination of the background error covariance matrix, fundamental for any assimilation scheme, is based on the innovation vectors (Berre and Desroziers, 2010) and is a difficult task because the true state is almost never known. In the available (wave) assimilation systems, approximations of the error covariances have been used.

In general, an arbitrary function of random input pairs χ and χ' is a valid covariance function when it is positive semi-definite (PSD). A real-valued, continuously differentiable function C is positive semi-definite in the neighborhood of the origin A if $C(0) = 0$ and $C(0) \geq 0$ for every non-zero $\chi \in A$. For covariance matrices, this definition is generalized for the corresponding self-adjoint matrix of the covariance function, which must be PSD. In other words, the covariance functions belong mainly to a specific relative small family of functions: constant, linear, polynomial, (squared) exponential, Matérn, rational quadratic and neural network (Rasmussen and Williams, 2006). In physical oceanography, the most common covariance functions are based on specific cases of generalized Gaussian distribution

$$\varphi(\chi) = \frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|\chi-\chi'|/\alpha)^\beta} \tag{1.2}$$

where χ *and* $\chi' \in (-\infty, \infty)$, $\alpha \in \mathbb{R}$ is the scaling factor, $\beta \in \mathbb{R}$ is the shaping factor and Γ denotes the gamma function. Bennett et al. (1996) used the product of Gaussian covariance functions for each dimension of the problem. Similarly, for the very first wave assimilation efforts, the empirical covariance function in physical space was determined as (Lionello et al., 1994)

$$C_{xx'} = V\exp(-|x - x'|/L) \tag{1.3}$$

where $C_{xx'}$ is the covariance between the two points x and x', V the variance and L is the optimal characteristic length. In this case L has been defined as five grid points (~1650 km). The comparison of the functions (1.2) and (1.3) show that (1.2) is a special case of (1.3) with $\beta = 1$ and $V = 1/2\alpha$.

In the wave assimilation system presented by Voorrips et al. (1997), the shape of correlation model, under the assumption that the background error is homogeneous and isotropic, was expressed as

$$C_{xx'} = \exp(-\left(\frac{r_d(xx')}{L}\right)^\beta) \tag{1.4}$$

where r_d is the distance between the points x and x', L the correlation length and β is a power to be determined. Both L and β were determined experimentally and equal to 200 km and 3/2 (as the best choice out of 1, 3/2 and 2) respectively. In other publications β was given different values. For instance Greenslade (2001) set $\beta = 1$ and Breivik and Reistad (1994) used $\beta = 2$, but with a decreasing L, from 200km to 40km during the minimization procedure. Breivik et al. (1998) determined L as a linear function of frequency with $L(0.04\,Hz) = 300km$ and $L(0.24\,Hz) = 60km$. Similar covariance functions have been applied in the temporal dimension of the problem. They are discussed in section 0.

The fundamental assumption of this research effort is that the covariance is calculated around the mean value of an ensemble of the quantity of interest (for instance wave spectral action density in time). The ensemble mean value is regarded as the best-guess estimate, while the ensemble spread is defined as a natural definition of the error variance (Evensen, 2009; Fisher, 2003). The covariance is expressed as

$$C_{NN'} = \overline{(N - \bar{N})(N' - \bar{N}')^T} \tag{1.5}$$

where C is the covariance and $N \equiv N(t, x, y, \theta, f)$ is the ensemble members in the space of the state vector, with x, y the spatial coordinates, θ, f the spectra coordinates and t the time. The overbar represents the ensemble mean and the superscript T denotes transpose of the matrix. The main issue with this approach is the computational cost, both for calculating the covariance matrices and applying them as weights during the assimilation.

Under this assumption, the approach to determine the covariance function in space (e.g. Eq. (1.3)) is expanded to all the dimensions of the wave spectral models (time, physical space, spectral space), and its applicability is investigated. The objective of the paper is to determine an appropriate background error covariance function for the weak constraint 4DVar assimilation system for nearshore spectral wave models. A main assumption in the development of the covariance function is that the covariance for each dimension is independent of each other and can be expressed as

$$C_f(x, y, \theta, f, t, x', y', \theta', f', t') = \prod V_i \, C_{xx'} C_{yy'} C_{\theta\theta'} C_{ff'} C_{tt'} \qquad (1.6)$$

where V_i is the variance of $i \in (x, y, \theta, f, t)$, defined as $V_i = C_{ii}$. Each factor of the product corresponds to each of the dimensions of the model. Each of the covariance function C of Eq. (1.6) will be determined based on its calculated covariance matrix.

The data used and the whole extent of the developed methodology are presented in section 2. The development of the background error covariance in time, in physical space and in spectral space are presented in sections 0, 0 and 3.3 respectively. Finally, in section 4, the main conclusions of this investigation are summarized and in section 5 the open questions on the subject are addressed as subjects of future research.

2   BUILDING THE BACKGROUND ERROR COVARIANCE FUNCTION FROM THE COVARIANCE MATRICES

The methodology to determine the covariance function is separated into two basic parts. The first part is the calculation and analysis of the covariance matrices based on measured wave spectra and simulated wave fields. The second part involves the synthesis of the covariance functions based on the covariance matrices by choosing and parameterizing an appropriate function in temporal, physical and spectral space. The procedure is depicted in Figure 1. The product of the resulting functions is the applied background error covariance function of the assimilation system. To determine the covariance function, a series of ensembles of SWAN simulations with different physical or numerical features of the model activated or deactivated was used. The resulting empirical function of the covariance is implemented in the SWAN-FAR assimilation system.

2.1   Wave Measurements

In order to determine the covariance on the basis of actual processes and model physics, all available wave measurements from the National Data Buoy Center (NDBC) of the National Oceanic and Atmospheric Administration (NOAA) for the period of three years, 2010-2012, have been analyzed (231 datasets in total). The geographical coverage of the data includes the N. Atlantic, Gulf of Mexico, N. Pacific (including Guam and Hawaii) and Great Lakes. The wave data have been used for two different purposes. The first is to determine the covariance function in time and the second as initial and boundary conditions of the wave field simulations.

2.1.1   Wave Data Preprocessing For Determination of Temporal Correlation Length

The calculation of the correlation ideally requires continuous time series, which is rare for (wave) measurements. Therefore, an hourly time series of a non-leap year is created by averaging the three measurements corresponding to the same date and time for the three different years. In case of missing values at the compiled time series, the gaps were filled by weighted interpolation. The weights depend on the number of the sequential missing values. In cases of time series with extensive (in order of days or more) missing values, the specific time series were excluded from analysis. A moving average filter of 12 hours is then applied to detrend the time series. The resulting series is analyzed both on a seasonal timescale, with three calendar months for each season, and also for extreme conditions defined here as having H_s more than 3m. The approach of the yearly averaging, in order to create the continuous time series, filters out the dependency on the instantaneous meteorological and oceanographic conditions, which has the added advantage of analyzing the dominant conditions of the area. The disadvantage of that approach is that characteristic lengths of short-term events (e.g. hurricanes) are not represented.
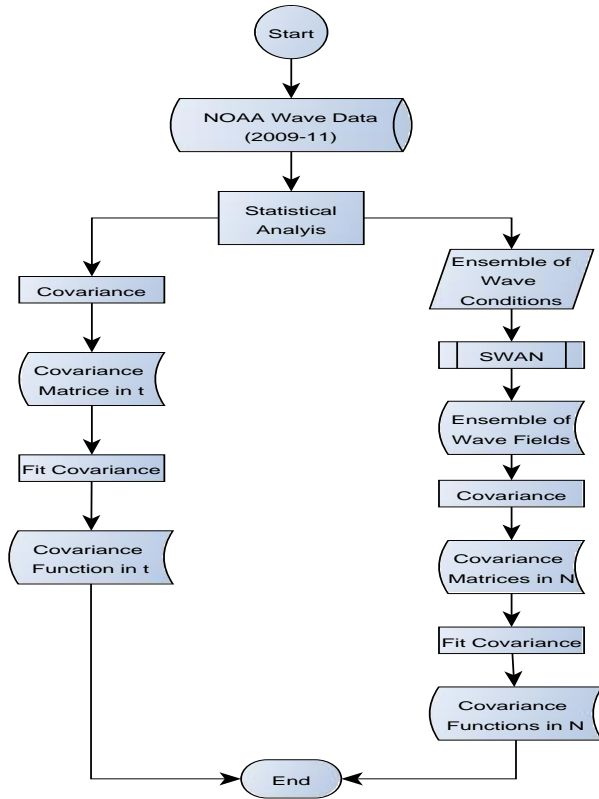
Figure 1. Flowchart of the steps to determine the covariance function based on calculated covariance matrices. *t* and *N* are time and spectral action density, respectively.

2.1.2 Wave Field Ensemble – Wave Climate Analysis

To determine the covariance matrices in physical and spectral space in coastal regions, simulated wave fields are analyzed. The simulations were forced with an ensemble of wave measurements generated by resampling the wave climate re-analysis. The 2011 time series from NDBC station 44014 (64 NM East of Virginia Beach, VA) was analyzed because it had a limited number of

missing values. The analysis was based on the calculation of the 3D joint-histogram of the significant wave height, peak period and corresponding direction denoted H_S, T_p, Dir_p, respectively. The joint probability distributions of the wave conditions were calculated according to the frequency of occurrence of each weather event (combinations of different wind/wave conditions).

The original time series was resampled by inverting the information of the corresponding histograms. In order to define the size of the new series, three criteria were applied: (1) to keep more than 80% of the original information, (2) to not have significant change at the calculated covariance matrices independently of the length of the ensemble, and (3) reasonable calculation requirements, in time and memory. Figure 2 shows the comparison of the statistics of original set (top row) to the reduced data set (bottom row).
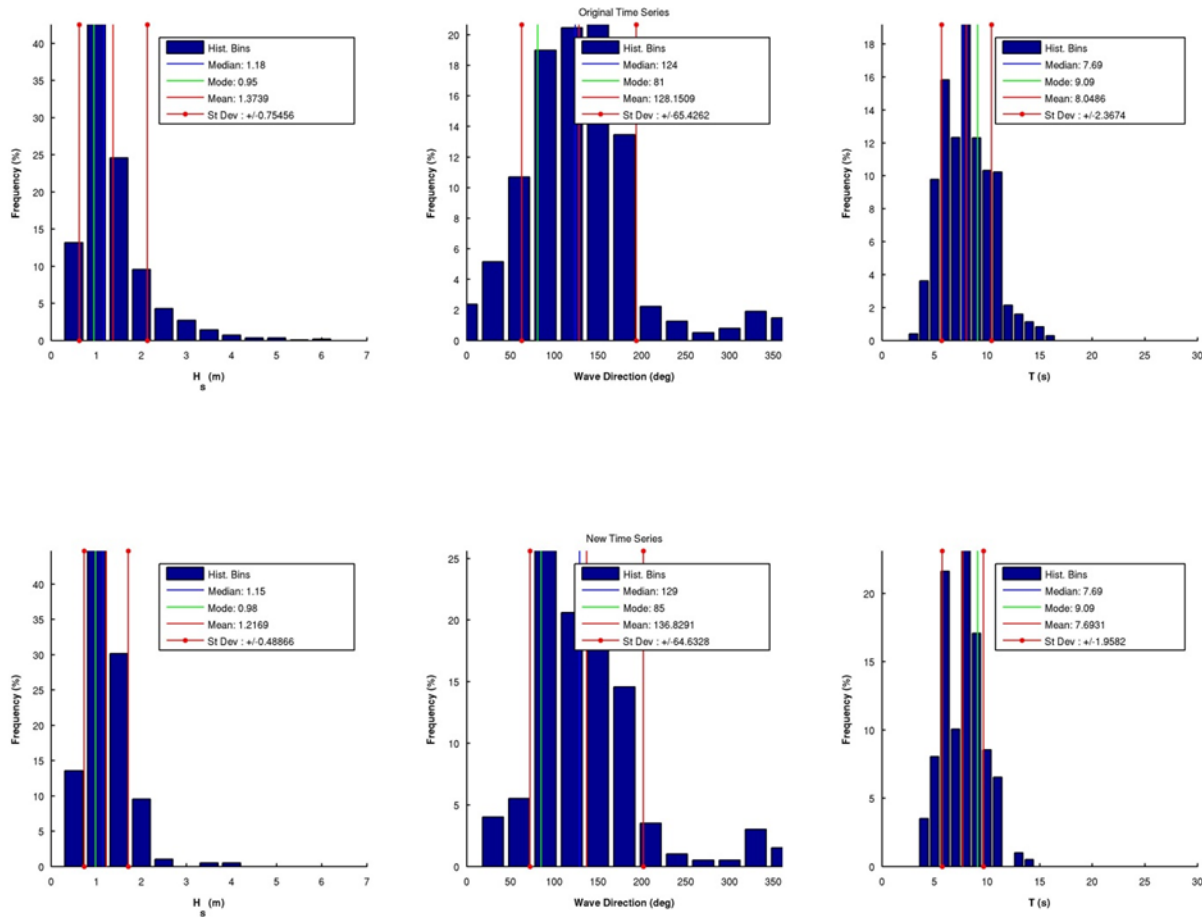
Figure 2. Top row: Statistical analysis of the wave-field properties at the station 44014 with 8000 samples. Bottom row: Statistical analysis of the resampled time-series with 200 samples.

2.2 Wave Model and Model Setup

The wave model used in this study is the third generation wave model SWAN, version 40.81 (Booij et al., 1999; Holthuijsen, 2007). SWAN solves the wave action balance equation

$$\frac{\partial N}{\partial t} + \nabla \cdot \left((\boldsymbol{c_g} + \mathbf{U})N \right) + \frac{\partial c_\sigma N}{\partial \sigma} + \frac{\partial c_\theta N}{\partial \theta} = \frac{S_{tot}}{\sigma}, \qquad (2.1)$$

where N is the spectral action density, $\boldsymbol{c_g}$ is the energy propagation velocity in the physical space, (c_σ, c_θ) the propagation velocities in the spectral space and S_{tot} represents the source terms that include wave generation by wind, bottom dissipation, surf-breaking, wave-wave interaction, bottom and current induced shoaling, and σ is the wave frequency.

2.2.1 Model Setup

Because there are five dimensions *(t,x,y,σ,θ)*, the determination of the covariance function is computationally challenging. Therefore, 15000 simple and idealized simulations are set up and analyzed. Due to the volume of the data, only results from the simulations with significant impact on the development of the covariance function are presented here. For all simulations, the default model physics and numerics are used, with the only exception being the use of the BSBT scheme instead of the GSE. The number of grid points in cross-shore direction, M_x, is 100; the number of frequencies, M_f, is 33, the frequency varies between 0.01Hz and 0.4775Hz; the directional resolutions, $\delta\theta$, is 10°.

Three different sets of bathymetric grids have been used in the current investigation:
  A. One dimensional with constant depth $d \in [1: 10, 20, 50, 100]$; d is in m.
  B. One dimensional with constant slope, $S_d \in (0.01, 0.1)$ dissipative and reflective, respectively, according to Komar, (1998), where S_d is the bottom slope.
  C. Two dimensional real case at the Field Research Facility Duck, NC. Dimensions of the grid are $M_x = 100$ grid cells and $M_y = 1000$, with $\delta x = \delta y = 20$m, where δx and δy are the cross-shore and longshore spatial resolutions, respectively. For the two dimensional simulations, the cross-shore direction corresponds to the *x*-axis and the long-shore direction corresponds to the *y*-axis.

In order to simplify the problem, instead of using the measured spectra as boundary conditions, the model simulations used two sets of parameterized spectra based on WAFO 2.5 (Brodtkorb et al., 2000). The first set consisted of unimodal spectra by assuming fully developed sea and applying the Pierson-Moskowitz function (Pierson and Moskowitz, 1964). The second set included swell, and assuming bimodal spectra, we used the approach described by Torsethaugen (1994) where the spectrum is assumed to be the sum of two modified JONSWAP spectra and the energy is divided between the two peaks according to empirical parameters. The directional spreading is based on Hasselmann (1973). The directional resolution is 10°, and the propagation direction is assumed perpendicular to the open boundary. In addition, model simulations were conducted with and without wind. In the wind forced runs, the measured wind conditions at the buoy location were imposed on the whole domain. The time-series of the forcing data are depicted in Figure 3.
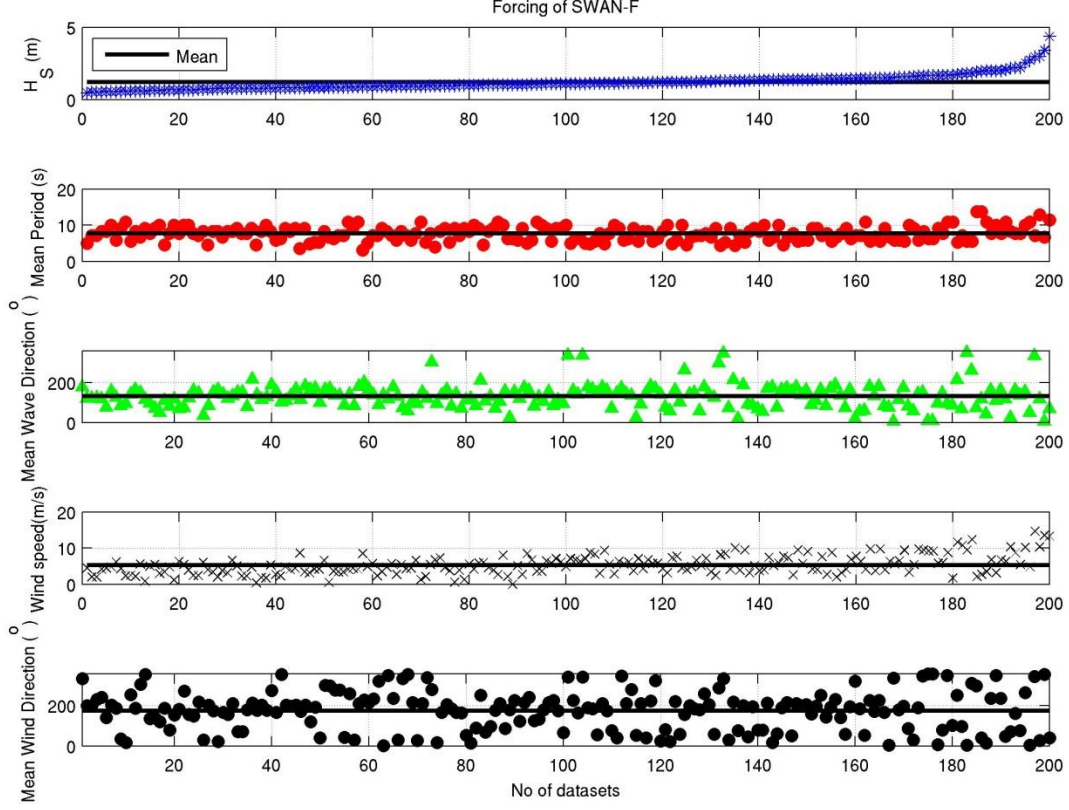
Figure 3 (Top to bottom). Time-series of H_s, mean period, mean wave direction, wind speed and mean wind direction, used as boundary conditions for the present investigation. The direction 180° corresponds to the direction towards the shore for both waves and wind. The black line in each of the plots represents the mean of the plotted quantity.

2.3   Calculation of Covariance and Correlation

As mentioned previously, it is assumed that the error covariance is independent in each dimension of the model. Hence, the covariance has to be calculated for each spectral bin, grid node, and time step. Rewriting eqn (1.5), the covariance for this study of two bins of spectral action density, $N(r)$ and $N'(r)$, of an ensemble $N^J(N_1, N_2, …, N_M)$ is calculated as

$$C(N, N') \approx \overline{((N - \overline{N})(N' - \overline{N'}))} = \frac{1}{M - 1} \sum_{i=1}^{M} (N_i - \overline{N})(N'_i - \overline{N'}) \qquad (2.2)$$

where M is the length of the ensemble. It is important to recall that N has five dimensions.

An example of the covariance matrix of two spectral bins of action density spectrum (quantity used at the SWAN-FAR) is depicted in Figure 4 (left).
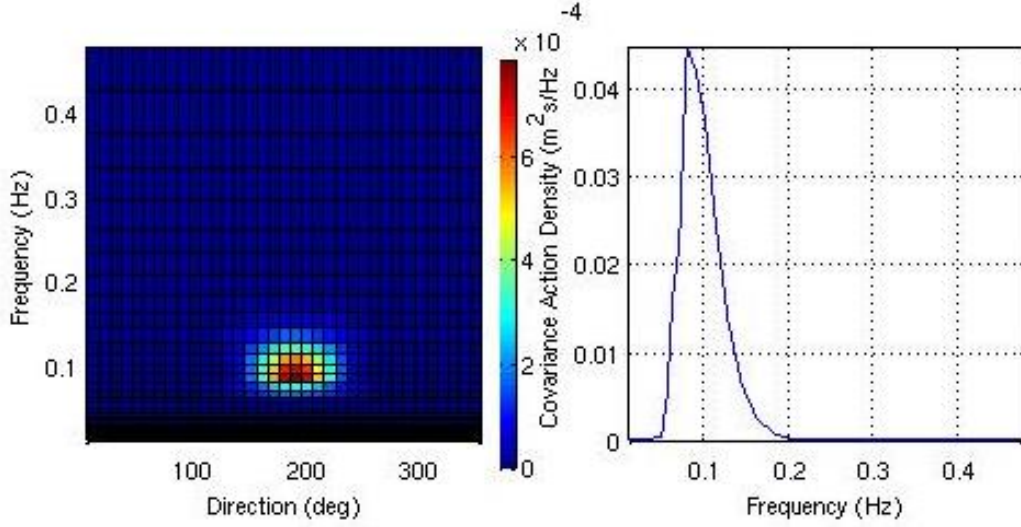
Figure 4. Left: Directional covariance spectrum of action density; Right: frequency spectrum of the error covariance for two spectral bins.

The covariance matrices have the same format as the SWAN spectra and based on this external characteristic, are referred to as Covariance Spectra. By the definition of the covariance, the units of the covariance spectra are the square units of the input spectra. In the next section, the integrated covariance spectrum over direction will be used for the development of the covariance function. An example is illustrated in Figure 4 (right).

Similarly, following the definition of the correlation, the correlation between the two spectral bins of the $\mathbf{N}$ and $\mathbf{N}'$ is defined as

$$R(\mathbf{N}, \mathbf{N}') = \frac{C(\mathbf{N}, \mathbf{N}')}{V(\mathbf{N})V(\mathbf{N}')} \tag{2.3}$$

where V is the sample variance of each dataset and C their covariance; the correlation is normalized covariance and both of the two quantities are used in the current investigation. For the rest of the analysis, ergodicity, stationarity and normality are assumed about the data.

An important limitation of the present analysis is the computational cost of the covariance. Indicatively, for each one-dimensional case, more than 100000 covariance matrices are calculated, and the covariance matrices calculation is expensive in terms of computational time and memory. For instance, for an ensemble of stationary simulations on a grid with $M_x \times M_y$ nodes (M_x and M_y are the number of nodes in x- and y- direction respectively) and output spectra for each node with M_f number of frequencies and M_θ number of directions, the covariance matrix of one impulse with all the other impulses includes $M_x \times M_y \times M_f \times M_\theta$ calculations and the total number of calculations is equal to $(M_x \times M_y \times M_f \times M_\theta)^2/2$. For one

ensemble of the one-dimensional cases described in the previous section, circa 7 billion covariance calculations take place. Apparently, for realistic cases the extreme number of calculations prevents the calculation of covariance matrices. Thus, whenever is possible, scientifically reasonable assumptions are made, in order to accelerate the estimation of the matrices. Hence, the covariance matrices are calculated for spectral bins with action density higher than $|O(10^{-12})|$.

The objective of this study is the determination of an optimized covariance function based on the covariance matrices. The first step is the analysis of the covariance matrices, in order to (1) identify their properties and (2) to reduce the data volume. The basic assumption for the analysis is the determination and modeling of the maximum impact that the covariance could have. Therefore, prior to any analysis, all the covariance matrices are normalized with the maximum value of the covariance in the dimension of analysis, e.g. in space, and based on the relative spectral density; the covariance spectra with low density are excluded from further analysis. Based on the properties of the remaining data, the appropriate theoretical functions are fitted to the data, and they are evaluated according to basic goodness of fit statistics.

In the following sections, the described methodology is applied on all the available data from both the measurements and the simulations. The covariance matrices are analyzed in the space of the model and the corresponding normalized covariance functions will be introduced.

3   COVARIANCE FUNCTION

The error covariance function is developed under the assumption that the best covariance error estimator is the covariance matrices around the mean of forward simulations or measurements. Hence, in the following paragraphs, measurements and simulations are treated in similar ways in order to analyze the properties of the covariance matrices for all the five model dimensions. The covariance function(s) is synthesized by expressing in mathematical terms the conclusions of the analysis. Due to the volume of the data and the computational cost, we optimize the analysis based on reasonable algorithmic simplifications and assumptions. In the following paragraphs procedures, analysis and synthesis, are extensively presented.

3.1   Covariance Function in Time

The temporal covariance function, $C_{tt'}$, between two moments, t and t', is defined as

$$C_{tt'} = V_t^{0.5} e^{-\frac{|t-t'|}{\tau}}$$ (3.1)

where $\tau \in \mathbb{R}^+$ is positive empirical constant and corresponds to the temporal correlation length and $t \ and \ t' \in [-\tau, \tau]$ correspond to two temporal instants of the analyzed data. This correlation usually is convolved with other fields by solving two Langverin equations (Bennett et al., 1996). The advantages of Eq. (3.1) are discussed by (Ngodock, 2005), and it is applied in most 4D-Var assimilation systems. A recent example is the 4D-Var system for the Navy Coastal Ocean Model (Ngodock and Carrier, 2014).

It has been proved that the temporal dependency in time of the wave measurements is determined via the calculation of the autocorrelation function of the property of interest, e.g.(Spanos, 1983), and based on this, the covariance function in time is determined by calculating the autocorrelation of the time series of significant wave height. The main assumption is that the maximum covariance in time is determined by the autocorrelation. From a technical point of view, it is assumed that the wave measuring devices and applied algorithms (different buoy manufacturers, ADCP) have similar biases and that the acquired time series of the wave amplitude have the same length and have been analyzed with the same algorithm, and therefore, all the wave measurements have the same accuracy and the same number of effective degrees of freedom. Moreover, it is assumed that autocorrelation of the energy of each direction-frequency spectral bin has the same properties in time with the autocorrelation of the integrated energy of each spectrum.

In general, the applied methodology follows the classical analysis of (Box and Jenkins, 1970) as applied on wave analysis from several researchers e.g. (Parvaresh et al., 2005; Soares et al., 1996; Yim et al., 2002). All these studies showed decreased of the autocorrelation of the H_s with the increase of the temporal lag, as expected. This common conclusion was drawn based on data from different locations with different wind and wave climatology, having different temporal lengths and acquired with different instruments. Intuitively, by observing the plotted data of publications on the subject, autocorrelation functions seems to follow specific functions. In some cases, the autocorrelation function is linear,

$$R(t) = -\frac{1}{\tau}|t| \tag{3.2}$$

where $\tau \in \mathbb{R}^+$ is the time scale and experimentally defined and $t \in \mathbb{R}^+$ and corresponds to the time lag. The main issue with function (3.2) is that it is not continuously differentiable, but still it could be used by defining , $t \in (0, \tau]$ and assuming symmetry.

In the cases of dissipating wavefields, the shape of autocorrelation function of time series of H_s follows normal distribution, therefore it could be represented as

$$C(t) = \sqrt{2\pi b_t}\exp(-[\frac{(t - a_t)}{b_t}]^2) \tag{3.3}$$

where $a_t \in \mathbb{R}$ and $b_t^2 \in \mathbb{R}^+$ both are empirically determined. In this study, the autocorrelation is calculated based on the function (2.3), using time series of H_s as

$$C(\tau_{Hs}^k) = 1/(M_t - 1 - k) \sum_{i=1}^{M_t-k} (H_s^i - \overline{H_s^i})(H_s^{i+k} - \overline{H_s^{i+k}}) \tag{3.4}$$

where $C(\tau_{Hs}^k)$ is the autocovariance function for $\tau_{Hs}^k = k\delta t$ to be the kth lag, with $k = 0,1, ..., M_t$ and H_s^i is the time series of H_s with length M_t.

From the definition of the characteristic correlation length, τ is calculated as

$$\tau = \int_0^\infty C(\tau_{Hs})d\tau_{Hs} \tag{3.5}$$

However, in practice Eq. (3.4) becomes

$$\tau = \int_0^{C_\tau^{lim}} C(\tau_{Hs})d\tau_{Hs} \tag{3.6}$$

where C_τ^{lim} is defined as the value of $C(\tau_{Hs}^k)$, when $C(\tau_{Hs}^k) \equiv 1.96/\sqrt{M_t}$, under the assumption of normality. Equation (3.6) is presented here for the time variable, but the correlation length can be similarly determined for any dimension of the problem.

Equations (3.1) and (3.3) express, in mathematical terms, that the autocorrelation function (ACF) in time of the wave integral properties follow (or almost follow) a normal distribution. Based on this, the normal distribution function is used for the estimation of the temporal covariance function. The first step in the analysis is the calculation of the normalized autocorrelation function of the linearly de-trended and temporally smoothed time series of the measurements of the significant wave height. The open issue is the selection of the appropriate temporal length of the input data for the estimation of the ACF so that the covariance properties of the simulated physical processes are represented. For instance, analyzing one year of hourly data, estimates a temporal correlation length characteristic of the wave climate of the specific area, but not for the short-term events (e.g. Hurricanes). Thus, two different timescales were selected:
   A. seasonal, by separating one year of data in quarters (JFM: January, February, March; AMJ: April, May, June; JAS: July, August, September; OND: October, November, December), and
   B. focused on extreme events by analyzing data with $H_s \geq 3m$.

Following NOAA, the available wave datasets are also sorted in seven subsets according to their location (ATL-SE: Atlantic Ocean Southern than Virginia; ATL-NE: Atlantic Ocean Northern than Virginia; GoM: Gulf of Mexico; GRT-LK: Great Lakes; PAC-CS: USA coast of Pacific Ocean; HAWAII and GUAM).

Figure 5 illustrates an example of the normalized autocorrelation of H_s time series for each season (asterisks with different colors), with the solid lines corresponding to the fitted normal distribution and the dashed line the critical value for determining the correlation length, τ, according to (3.6). The analyzed data is from NDBC buoy 44005, located offshore the North East US coast at 200 m depth. Due to its location, there is minimal impact from nearshore processes and human activities on the wave record. The question to be answered is if the function (3.3) is the appropriate covariance function to model the temporal covariance. After the fit, the calculated values of the ACF are tested to determine if the normal distribution-like function with the appropriate goodness-of-fit measures is applicable. There are several reasons to reject the hypothesis of normal distribution, including buoy deployment location, wrong choice of buoy size, tide influence on the data which was not filtered out during pre-processing, and artifacts due to the compilation of one continuous time series by combining data from different

years. The datasets which fail to be modeled with the normal distribution are excluded from the synthesis of the covariance function without further investigation. Equation (3.3) is fitted to the remaining datasets. In Figure 5, the parameters a_t and b_t for the specific buoy are listed. In the same figure, the plots of the fitted functions strongly indicate that (3.3) can be used as covariance function. For the evaluation of the fitted functions, the $R^2 > 0.90$ and the root mean square error are used.
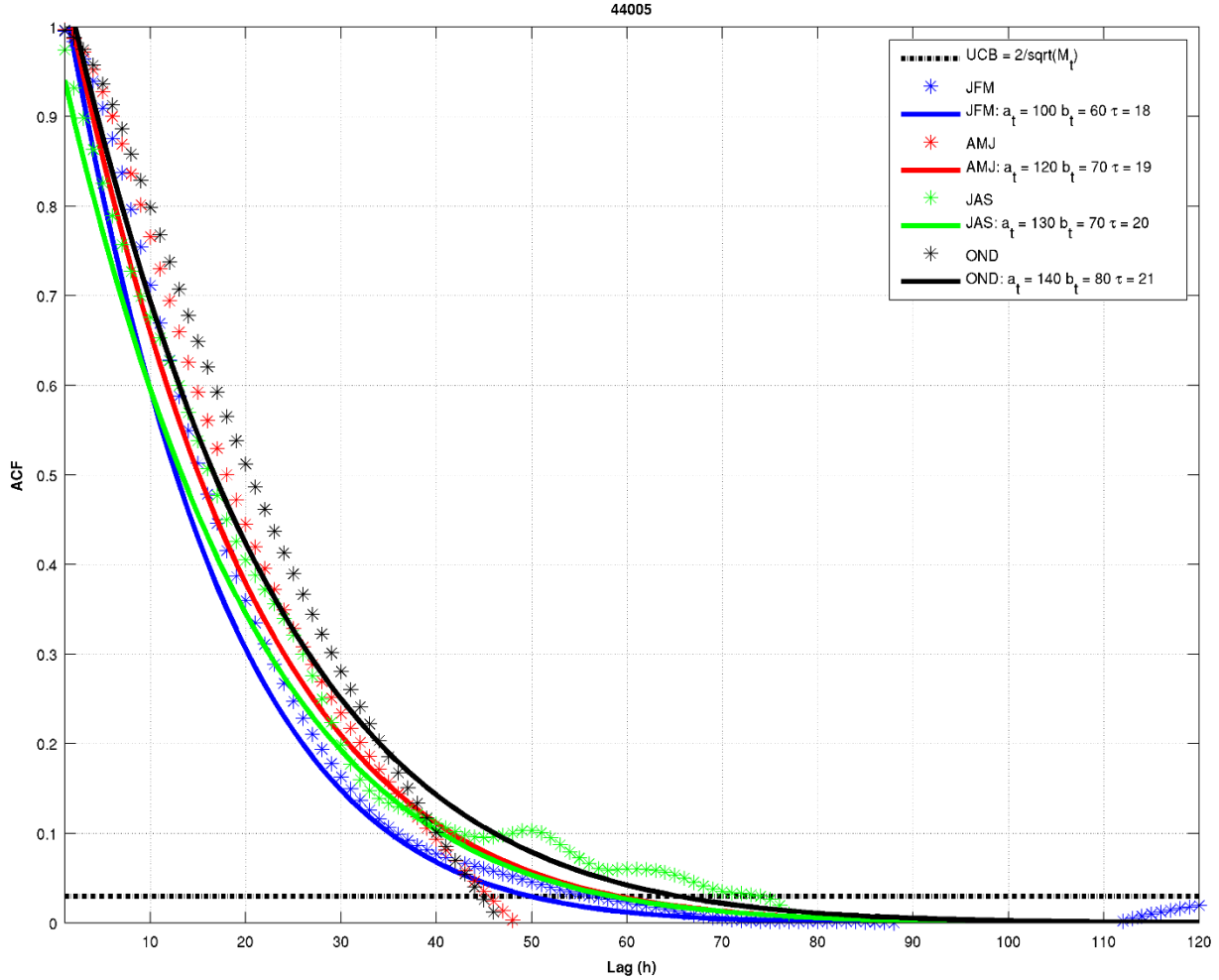
Figure 5. The ACF of H$_s$ measured at the 44005 NDBC buoy for the four quarters of the time series. The four colors correspond to the four quarters. The asterisks represent the calculated ACF and the continuous lines are the results of the fitted function (3.3). The fitting parameters are listed in the figure legend. The black dashed line, parallel to the x-axis is an indicative significance level, used for the determination of the correlation length for JFM.

Examining the ratio, $r_t = b_t/a_t$ between the two parameters, there is a clear seasonal and location dependency as seen in Figure 6. The mean ratio varies between -1.90 and -1.70 with standard deviation less than 0.1. This allows simplification of (3.3) and reduction of the computational time of the assimilation. As is seen in Figure 6, areas closer to the equator, especially Hawaii and Guam, and to a lesser extent GoM have two seasons, JFM-OND and

AMJ-JAS, as expected. During the JFM-OND, the ratio varies around -1.8 and during the AMJ-JAS, the ratio is reduced to lower than -1.85.
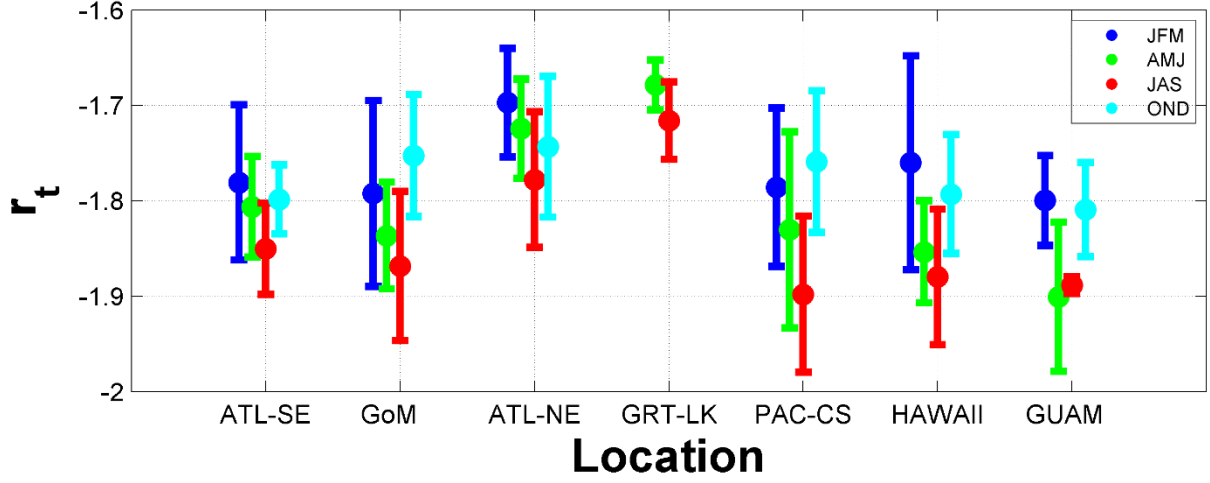
Figure 6. The mean and standard deviation of the ratio r_t calculated for the seven areas defined by NOAA.

The Atlantic coast of N. America is separated into ATL-SE and ATL-NE, South-East and North-East USA coasts, respectively. ATL-SE has similar values as the three equatorial areas, but based on the ratio the winter period is longer, extended from October to June; the ratio is almost constant and approximately -1.8. During summer (JAS), the ratio value is reduced to -1.85. In ATL-NE, the ratio reaches its highest value during winter (JFM), -1.7. During spring (AMJ) and fall (OND), the ratio is almost the same and a little higher than -1.75. During JAS, $\overline{r_t}$ has the minimum value for the specific area. In general, the ratio values illustrate the differences in wave climate between the N. Atlantic and almost equatorial Atlantic Ocean. Along the Pacific coast, PAC-CS, there are clearly four seasons according to the values of $\overline{r_t}$. From JFM to JAS, a gradual reduction of the ratio is observed (from -1.8 to -1.9), but its peak value, -1.75, is reached during OND. During AMJ-JAS, the PAC-CS and Hawaii have very similar values, but during the fall (OND) and winter (JFM), the two areas are influenced from wave-fields originating from different locations. The highest values of the ratio $\overline{r_t}$, are seen in the Great Lakes. There are a large number of missing values during OND and JFM. The similar values of the ratio for both AMJ and JAS and small standard deviation of the $\overline{r_t}$ show that the wave conditions are dominated by local conditions, for instance the orientation of the lakes in comparison with the wind direction.

3.1.1 Climatological Temporal Correlation Length

The missing information are the actual values of the temporal correlation length, $\overline{\tau}$, which are necessary for the application of any of the two suggested covariance functions (3.1) and (3.3), τ and b_t, respectively. Hence, in order to generalize the output, the procedure for the results of Figure 5 is applied to all the available time series. The spatial distribution of the correlation lengths is shown in Figure 7. The temporal correlation lengths have significant regional differences, and the need to calculate the statistical properties and determine the statistical

14

relationships at regional or even local scales is evident from the figure. By using the same spatial clustering as in the previous section, the median for each region and season of the correlation length are calculated and shown in Figure 8.
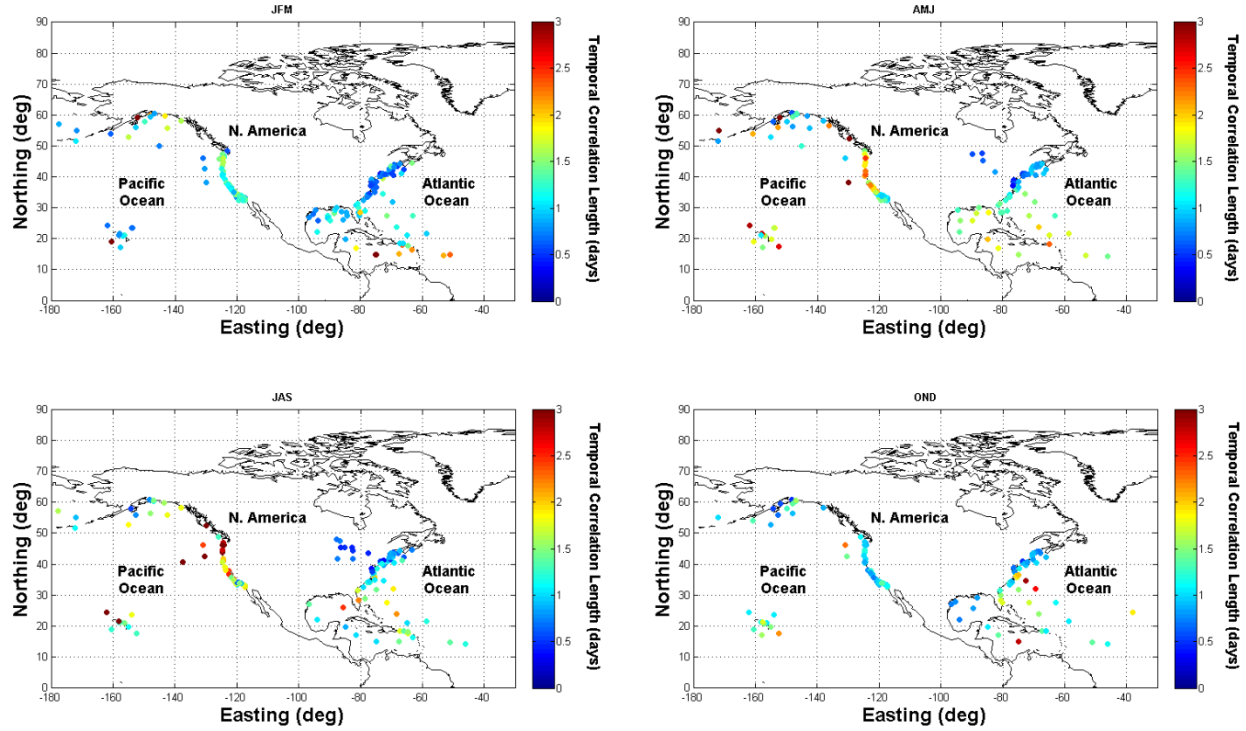
Figure 7. Seasonal distribution for the temporal correlation length for N. America based on all the available NOAA measurements for the period 2010-2012.

The mean temporal correlation length for all three Pacific Ocean areas (PAC-CS, Hawaii and Guam) are very similar. The relatively constant conditions due to the long waves are probably the reason for the long temporal correlation length at the Pacific coasts of USA and Canada. During the combined period of JFM-OND, $\bar{\tau}$ varies between 1.2 and 1.5 days for all the three areas. The maximum values for Hawaii and Guam (which are located almost at the same latitude) are observed during AMJ and is almost two days, and the maximum value of temporal correlation for the PAC-CS is during JAS. At this point, it has to be emphasized that the calculated mean values for PAC-CS cover the area between California and Alaska. Thus, for regional or local experiments, the actual values displayed in Figure 8 should be used. The spatial distribution of the values of $\bar{\tau}$ shows that the maximum values are along the coasts of Washington and of British Columbia during AMJ-JAS, and the values are higher than 2 days.

In the Atlantic Ocean, there is a remarkable difference between North (ATL-NE) and South (ATL-SE). Though the values of $\bar{\tau}$ in both regions are almost constant during the year, the values in ATL-SE are approximately 1.5 days, which are much higher than in ATL-NE, where the temporal correlation length is 0.8 days. Despite the spatial and temporal consistency of the correlation length values, there are specific locations that have permanently higher values than the average, such as in the Florida straits and along the east Florida coastline. It is likely that here the permanent, strong Gulf Stream dominate the wave field characteristics.
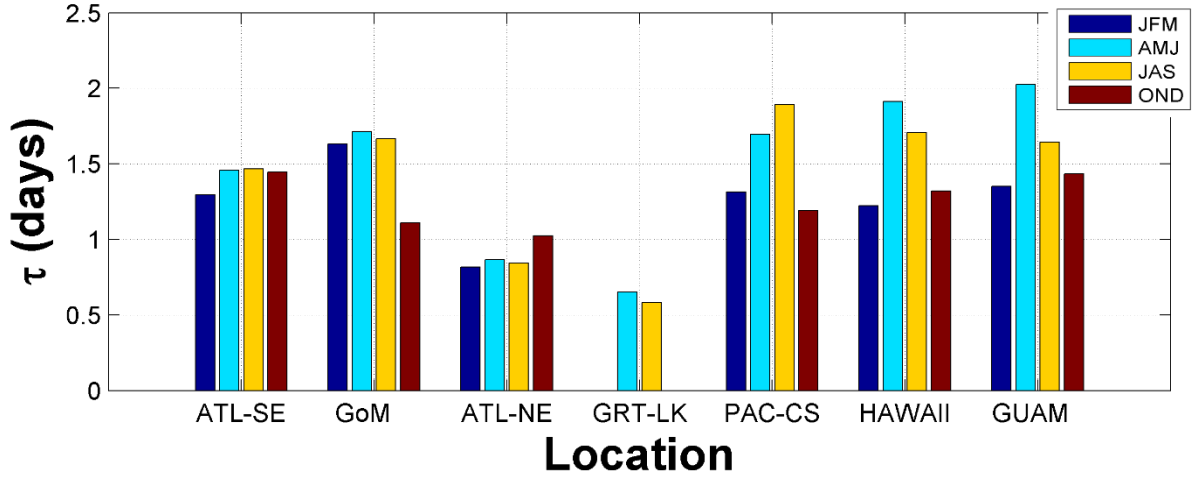
Figure 8. Synopsis of the spatial information in Figure 5: Median of seasonal correlation length for the seven different areas that there are wave measurements by NOAA for the period 2010-2012.

By extending this interpretation in the Gulf of Mexico, where the circulation is dominated by the consistent loop current, the values of up to 1.5 days correlation length seems reasonable. In addition to that, the GoM is semi-closed with almost constant wave climate and a limited number of extreme events (e.g. storms, hurricanes) per year. It has to be mentioned that during the three years of the analysis, there was only one category 1 storm (Hurricane Isaac in 2012) smoothed during the data preprocessing. The Great Lakes have approximately 0.5 days correlation length in time. The limiting factor, in this case, is the fetch and the influence of the local wind fields. Similar temporal correlation lengths have been calculated for seas with similar geographical characteristics, for instance the Aegean Sea (Flampouris, 2003).

3.1.2   Temporal Correlation Length of Extreme Events

Extreme wave events are defined as periods of 48 h or longer during which the significant wave height is higher than 3 m. To get the correlation length during such events, the wave data from all the three years at all locations of measurements have been analyzed by following the same methodology as for the climatological analysis. The temporal correlation lengths are given in Figure 9, where the values of τ have been plotted as a function of the location (buoy). The colors correspond to the same seven areas as in the previous section. The filled circles are the mean values of τ for each buoy, and the bars reveal the minimum and maximum value of τ. The black line is the mean of the plotted means for each area.
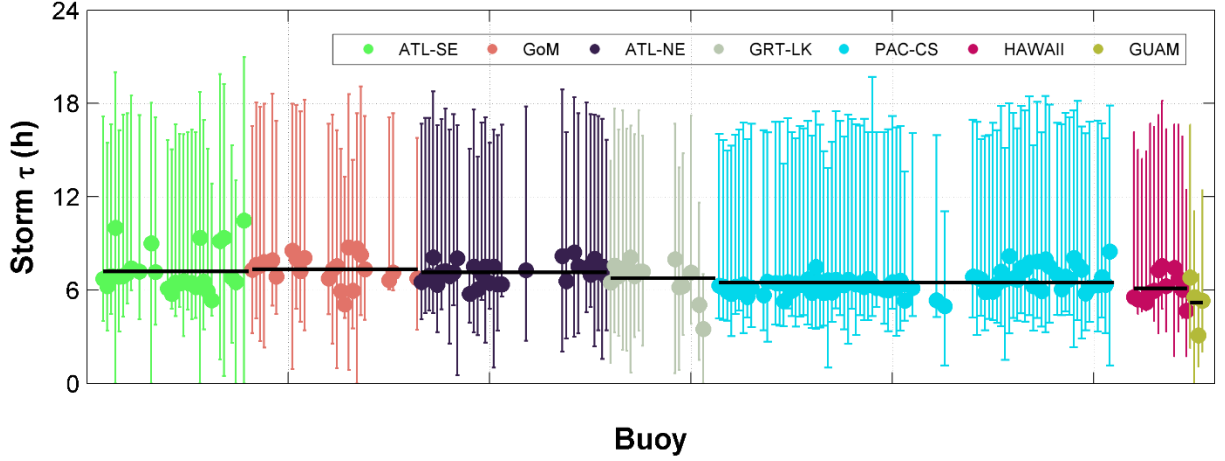
Figure 9. Temporal correlation length for cases with $H_s \geq 3m$ for each available buoy. The dots correspond to the mean temporal correlation length and the lines the minimum and maximum value for each buoy. The black lines show the mean of the correlation length for each of the seven domains.

As is evident from Figure 9, the mean τ is almost constant and for all the buoys varies between 6 and 7 hours. The difference between the mean and minimum values is approximately 2 hours, but the maximum values are almost three times higher than the mean, 18 h. The temporal correlation length is independent of the measurement location in the case of storm conditions.

Summarizing the results, the analysis of long time series reveals the influence of the local conditions on the temporal correlation length. For extreme wave conditions of short duration, the statistics of the correlation lengths are approximately constant and similar independent of the location. The extended study of the temporal correlation lengths as a function of the external, dominant conditions is out of the context of the present investigation. The significance of the selection of the appropriate time scale for each phenomenon is the focus for the current section.

3.2   Error Covariance Function in Physical Space

The objective of this section is to determine the covariance function in physical space with a focus on nearshore applications. Even though sea surface wave fields are two dimensional, the covariance is a three-dimensional function of location and depth. For this part of the study, the covariance matrices of simulated wave fields are calculated, analyzed and modeled. To reduce computational cost and data volume, the following simulations are primarily conducted in a simplified one-dimensional format.

Very little has been published about the spatial covariance function for nearshore wave modeling applications. In the majority of published work on wave assimilation, spatial correlation lengths are on the order of hundreds of kilometers, for two reasons: (1) the coarse spatial resolution of the wave models for ocean simulations, and (2) assimilation of satellite wave data, which by default are sparse and are calculated over large areas. The present study focuses on a nearshore wave assimilation system for SWAN. Consequently, the spatial resolution of simulated wave fields is $O(10)$ meters, several orders of magnitude smaller than the published spatial correlation lengths.

The complexity and strong non-linearities in the nearshore waves (e.g. wave breaking, wave-wave and wave-current interactions) are significant obstacles to the development of a theoretical function for the covariance. Lacking published data on the spatial covariance of action density for shoaling and breaking wave fields, the spatial behavior of the associated covariance matrices is unknown. To construct a preliminary spatial covariance dataset for nearshore action density, an ensemble of 200 realistic wave field stationary simulations was created using bathymetry and data from the FRF in Duck, NC, and the covariance matrices were calculated for several grid points.
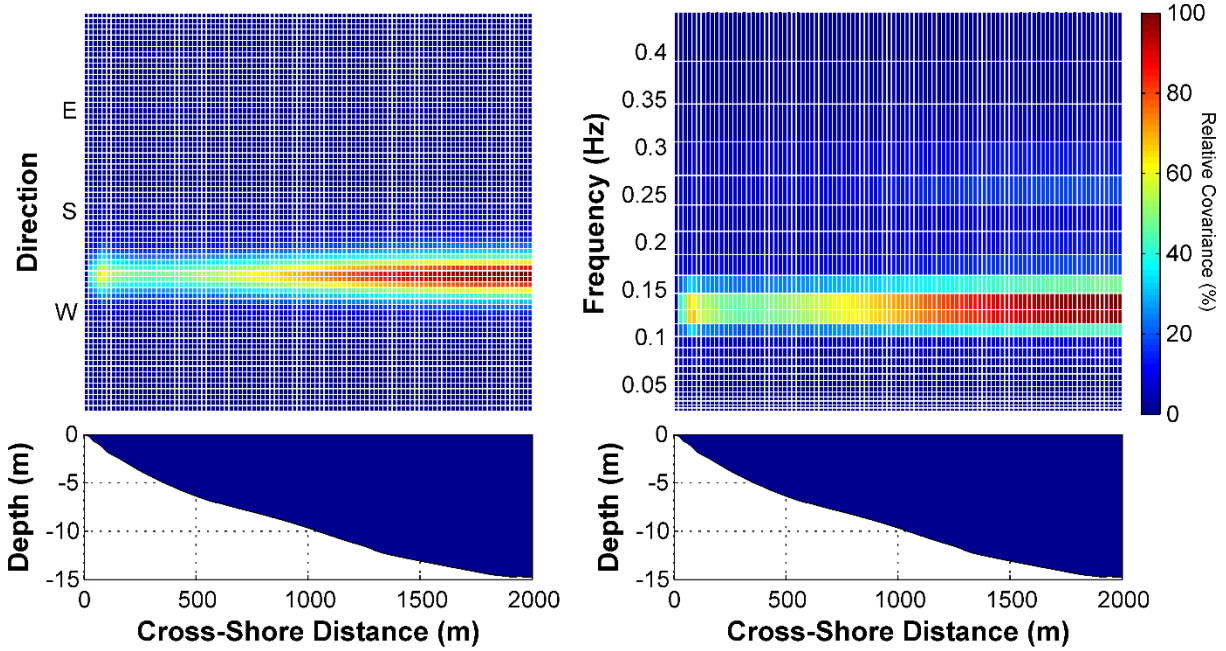
Figure 10. Alongshore-integrated and normalized covariance at Duck, NC as a function of distance and direction (left) and distance and frequency (right). The covariance was calculated between the point (1360 m, 840 m) and all other grid points.

In order to visualize and analyze the four dimensional covariance, the information of the covariance matrices was summarized by averaging and normalizing for each dimension. As the cross-shore (x-axis) bathymetry does not vary significantly over the long-shore direction (y-axis), the covariance spectra over the long-shore direction were averaged, thus the dimensions of the problem have been reduced to three: cross-shore, frequency and direction. In order to study the directionality, the covariance spectra were integrated over frequency and normalized by the maximum covariance value, hence x- and direction are the two dimensions of the covariance, Figure 10 (left panel). Likewise, the integrated covariance as a function of x- and frequency is examined, by integrating over direction and normalizing with the maximum covariance value, Figure 10 (right panel). The content of Figure 10 shows clearly the main characteristics of the covariance matrices to be investigated in the following sections. (1) The maximum covariance is a function of the depth. (2) The direction of the maximum covariance is normal to the shoreline, and its directional spread can be defined. (3) The maximum value of the covariance appears at specific frequencies. Based on these observations, the covariance function is examined by

treating each dimension of the problem independently. The covariance in the spectral domain is discussed in section 3.3.

3.2.1 Covariance as a Function of Distance

In the introduction, the most commonly used spatial covariance function (1.2) was presented. This general form of the equation, which was proposed as a solution of the diffusion-based correlation models, is the only covariance function in the literature being used for wave assimilation systems. Physically, it declares that the covariances are isotropic and homogeneous in a circular region centered on the grid point of interest with radius equal to the spatial correlation length *L*. Outside of the circle, the covariance is constant. In all published applications, as also mentioned earlier, the applied correlation lengths vary between 40km and 200km, which are several orders of magnitude higher than the spatial resolution of the computational grid for nearshore wave simulations.

To further simplify the problem and to separate the impact of the distance from the impact of the bathymetry, covariance matrices are calculated using one-dimensional grids of uniform depth. Each wave field ensemble is assigned a different fixed depth value, ranging from 1—$100m$. The covariance density is then integrated and normalized by its maximum value in the *x*-dimension; results are shown in Figure 11 (top). The maximum value of the covariance is at the boundary grid cell. This result is the same for all the ensembles since the same boundary conditions have been imposed for all cases. For each depth, the covariance tends asymptotically to a constant value that depends on the depth, the lowest of which occurs at $1m$ depth. The covariance is constant for depths higher than $20m$ (In the figure, the three lines corresponding to the covariance of depths higher than $20m$ are plotted on top of each other). At the boundary, the high value of the covariance is due to depth-induced breaking. To examine the change of covariance with distance, the gradient of the relative covariance in space has been calculated for each depth (Figure 11, bottom). The gradient for depths higher than $20m$ is equal to zero; thus, it is impossible to plot on a logarithmic scale. This plot confirms the impression given by the top panel that the covariance is almost constant. With the exception of the first few grid cells, less than 10 for all the depths, the order of magnitude of the gradient is 10^{-5}, or essentially zero. This extremely small reduction of the covariance can be explained as the dissipation of the wave energy due to the bottom friction. Due to the short distance, the effect of wave generation due to wind is negligible.
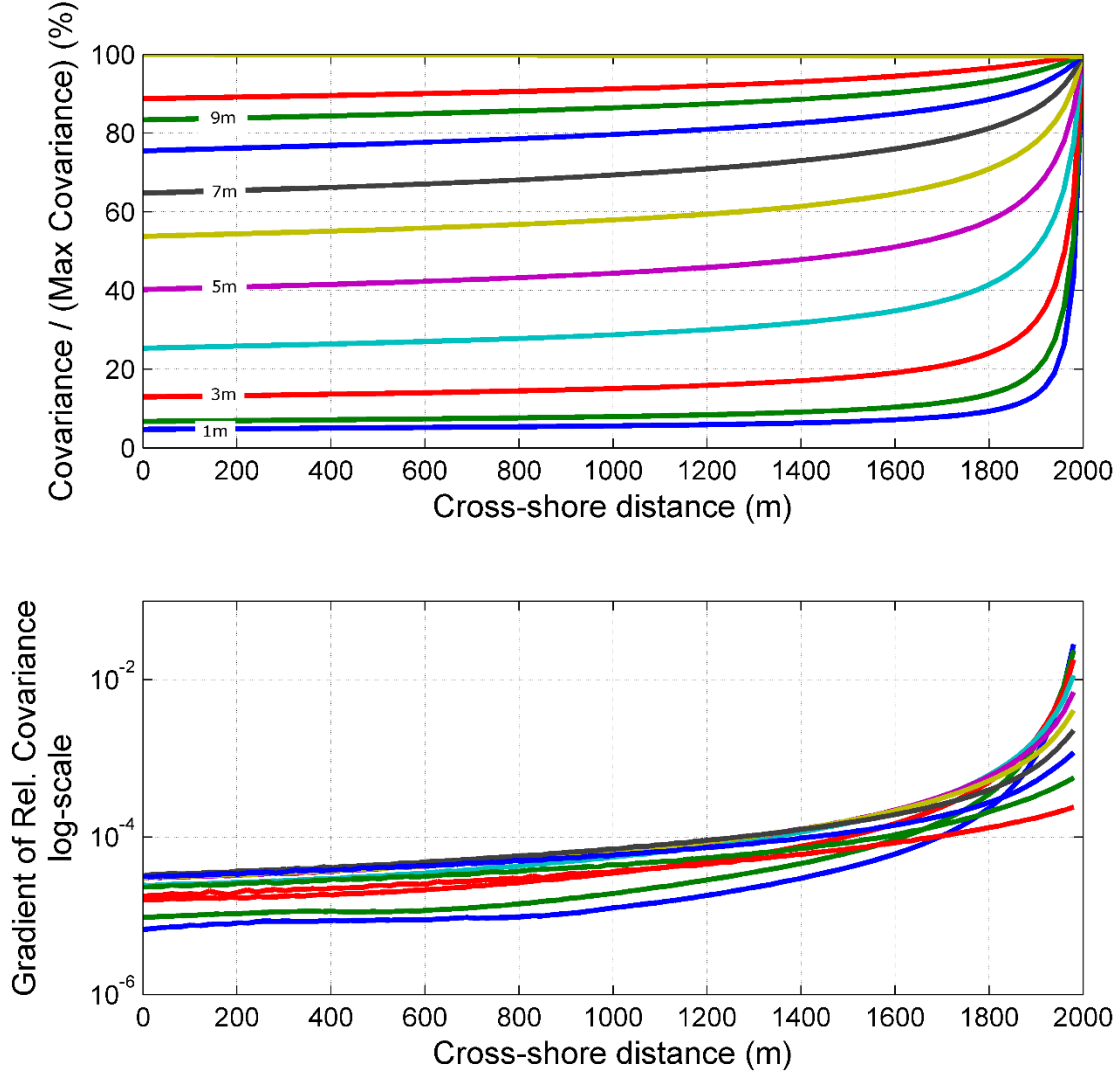
Figure 11. Top: The relative covariance as a function of distance for 13 different depths [1:10, 20, 50, 100] m and Bottom: its spatial gradient. The wave field propagates from the right to left.

For large scale wave simulations, the covariance in physical space is given by the expression in (1.4). In the following analysis, the *x*-direction will be defined as the normal to the local depth isoline and the *y*-direction as the tangent to the local isoline. Based on the results of Figure 10 and Figure 11, it is clear that the covariance for nearshore applications is a function of bottom-induced wave breaking and of the actual depth. Therefore, the function (1.4) has to be updated accordingly. Since the values of the error covariance can be considered constant in the *y*-direction and depending mainly on depth, the covariance function in space may be expressed independently for the two directions.

$$C_{xx'} = C_d[C_{xx'} + C_{yy'}] \tag{3.7}$$

20

where $C_{xx'}$, $C_{yy'}$ and C_d are the covariance functions in *x*- and *y*-direction and depth, respectively. If depth is a constant, we assume that homogeneity and isotropy holds, and therefore express the covariance function as:

$$C_{xx'} + C_{yy'} = \exp\left(-\frac{|x-x'|}{L_x}\right)^{\beta_x} + \exp\left(-\frac{|y-y'|}{L_y}\right)^{\beta_y} \tag{3.8}$$

where L_x and L_y are the correlation lengths for each axis and β_x and β_y are empirically defined parameters in each direction with suggested values between 0.5 and 3. The correlation lengths depend on the local conditions and dominant phenomena. From Figure 11, it is clear that L_y is on the order of magnitude of the long-shore direction of the grid. L_x is a function of the depth and the nearshore wave field transformations, mainly breaking. Therefore, L_x can vary between a few hundred meters (or grid nodes) and several kilometers. This issue is discussed extensively in the next paragraphs.

3.2.2   Error Covariance as a Function of Depth

In order to define the effect of the depth on the covariance of the wave field, the covariance matrices were calculated and analyzed for thirteen ensembles of one-dimensional forward simulations, each with constant depth. To reduce the volume of data and focus on the maximum effect of the covariance for the simulated wave field, a data subset is constructed by extracting the maximum error covariance values for each grid node. For this section, the covariance peak at the boundary is ignored and the analysis focuses on the grid points that are minimally influenced by the boundary effect. Figure 12 depicts the maximum normalized covariance for the whole length of the grid as a function of the depth. As the only variable in the forward simulations is the depth, it is apparent that the maximum value of covariance is calculated at the boundary and the covariance tends to a specific value, which depends on the actual depth.

In principle the behavior of the covariance near the boundary depends on the forcing and the traveling distance from the initial grid node at a depth that affects the wave field properties. In order to exclude these nodes, the second derivative in space of the normalized covariance was calculated, and the data were filtered according to it. The basic idea of using the second derivative is to determine the point that the first derivative becomes constant, implying that the reduction of the covariance is due to a permanent process such as bottom friction. The order of magnitude of the second derivatives of the covariance varies between 10^{-4} and 10^{-10}. Based on a calculation of statistical properties of the relative normalized covariance, the threshold for filtering was defined equal to 5×10^{-6}, as a mean conservative value. The number of excluded grid points depends on the depth, but in the majority of the cases, approximately twenty points are removed.
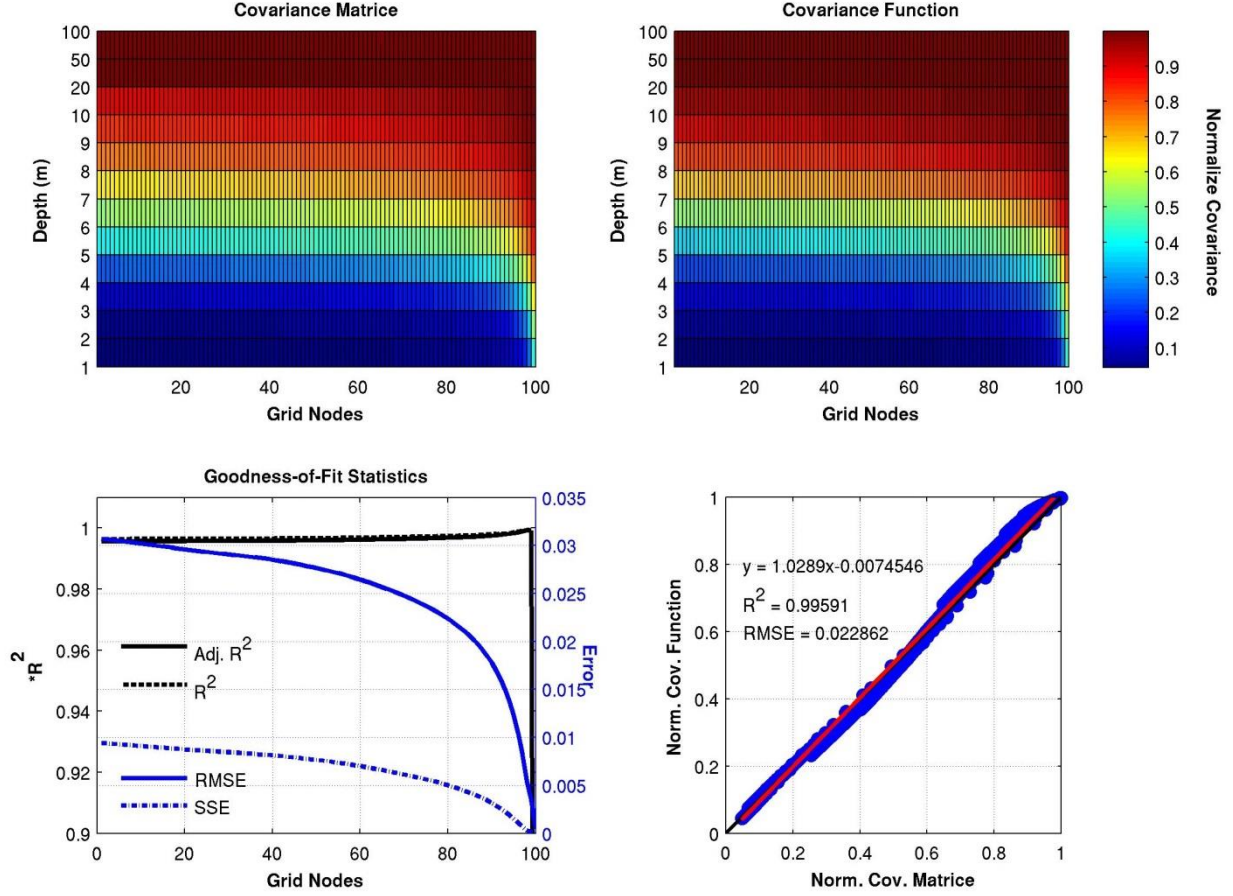
Figure 12. Top left (a): Calculated normalized covariance as a function of the depth. Top right (b): Estimated normalized covariance as a function of the depth. In both cases, the boundary is located at the grid node 100. Bottom left (c): Goodness-of-fit statistics of the covariance function for the entire computational grid. Bottom right (d): Scatter plot between the calculated covariance and the estimated covariance.

Based on these experimental results, the covariance function for depth can be described by an exponential expression of this form:

$$C_d = \alpha_d \exp(-(d - \beta_d)/\gamma_d)^2 \tag{3.9}$$

where $\alpha_d, \beta_d, \gamma_d$ are fit-defined constants and d is the actual depth. Equation (3.9) is a function of only depth and was fitted to each grid node individually. The results of covariance functions for each node are presented in Figure 12b. The estimated covariances have a high correlation with the simulation-based covariances (Figure 12a), $R^2 = 0.995$. The goodness-of-fit statistics, Figure 12c, confirm the accuracy of the covariance functions, with $R^2 \sim 1$ at all locations except the grid points close to the boundary. Similarly, the scatter plot between the two quantities, Figure 12(d), strongly confirms that a function in the form of (3.9) is the appropriate for modeling the covariance as a function of depth. The sum of the squares due to the error (SSE) and the root mean square error (RMSE) show the boundary effect on the covariances. Indicative

values of the constants are tabulated in **Error! Not a valid bookmark self-reference.** and the results are shown in Figure 13.

Table 1. Properties of covariance function according to the depth for different grid points and the mean from 1ˢᵗ to 80ᵗʰ grid point.

		Grid point									
		1ˢᵗ	25ᵗʰ	50ᵗʰ	75ᵗʰ	80ᵗʰ	90ᵗʰ	95ᵗʰ	99ᵗʰ	100ᵗʰ	Mean
Parameter	a_d	-1.65	-1.59	-1.52	-1.41	-1.37	-1.24	-1.10	-0.48	<0.001	-1.54
	β_d	-1.51	-1.35	-1.18	-1.03	-0.99	-0.97	-1.19	-1.42	60.38	-1.25
	γ_d	5.62	5.47	5.28	5.01	4.93	4.73	4.66	4.58	38.44	5.33

In order to determine a single representative function, each of the three parameters of (3.9) is averaged in space, excluding the boundary/breaking zone. Thus, the following empirical equation is derived for the normalized covariance as a function of the depth:

$$C_d = 10\text{^}[-1.54\,exp(-(d+1.25)/5.33)^2] \tag{3.10}$$

Figure 13 presents a comparison of the calculated covariance with the results of the covariance function for the specific grid node and with the results of the "average" covariance function (3.10). It is clear that the covariance function could be determined by fitting (3.9), which depends on only the local depth. Additionally, the difference of the "average" covariance function, (3.10), from the calculated covariance of the ensemble and the individual result of the individual covariance functions is negligible at locations far from the boundary; see e.g. the first nine plots of Figure 13. As the "average" covariance function is a mean representative function, it overestimates the covariance at the first grid points, mainly for depths higher than 5*m* and underestimates it for grid points close to the boundary.
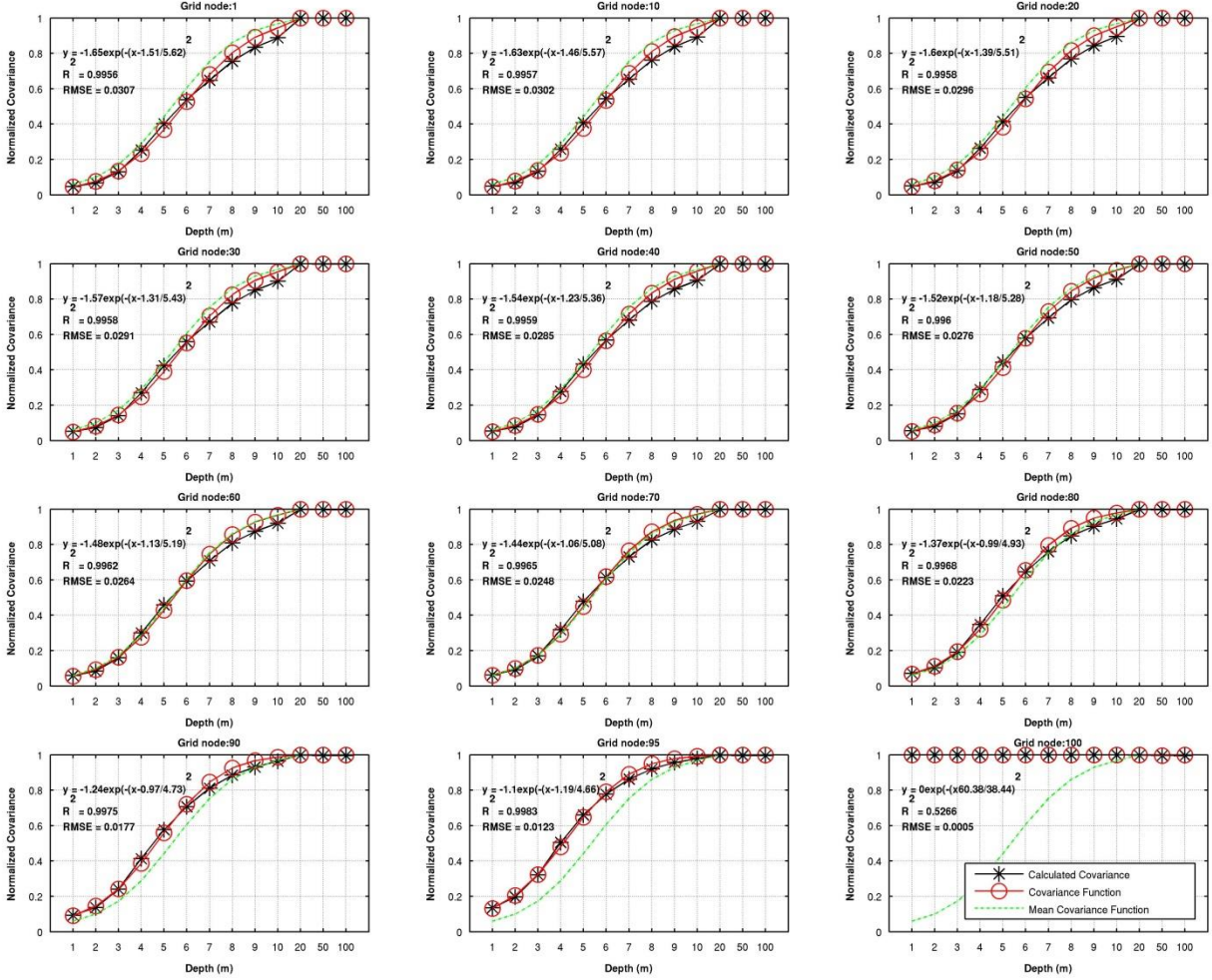
Figure 13. Characteristic examples of the maximum error covariances as a function of the depth for different distances (in number of grid nodes) from the boundary. The black line corresponds to the simulation covariance; the red line is calculated based on the covariance function for the specific grid node, and the green line corresponds to the mean covariance function.

3.2.3 Error Covariance of Breaking Wave Field

This section is focused on the determination of the covariance function of wave field breaking due to the local bathymetry. In Figure 10, at cross-shore distance $100m$, there is a peak in the integrated covariance. The location indicates that this peak is due to the local depth, or more precisely, onset of wave breaking when the waves reach the breaking depth. Mathematically, wave breaking is presumed to occur when the ratio $\gamma = H/D$, where H is the wave height at total depth, D, which includes the wave-induced setup, exceeds a prescribed limit. A typical value used in many simulations is $\gamma = 0.73$, which is also adopted for the simulations of this study.

As described in Section 2, the simulations are forced with climatological data acquired at $17m$ depth and are not adjusted to the depths of the simulation grids. The imposed wave boundary

24

conditions are dissipated by depth-induced breaking when $\gamma \geq 0.73$. Hence, the dissipation due to the depth is reproduced without taking into consideration the rest of the dissipation mechanisms. Following the same methodology as in the previous sections, the covariance matrices are calculated, and the covariance function is determined from them. Initially, the zone of significant variation of the covariance due to breaking is determined with the same criterion as in section 0. The exact lengths of the artificial breaking zones as a function of the depth are depicted in Figure 14 (left). By integrating the normalized second derivative in space, the correlation lengths for each depth are determined (Figure 14 right).
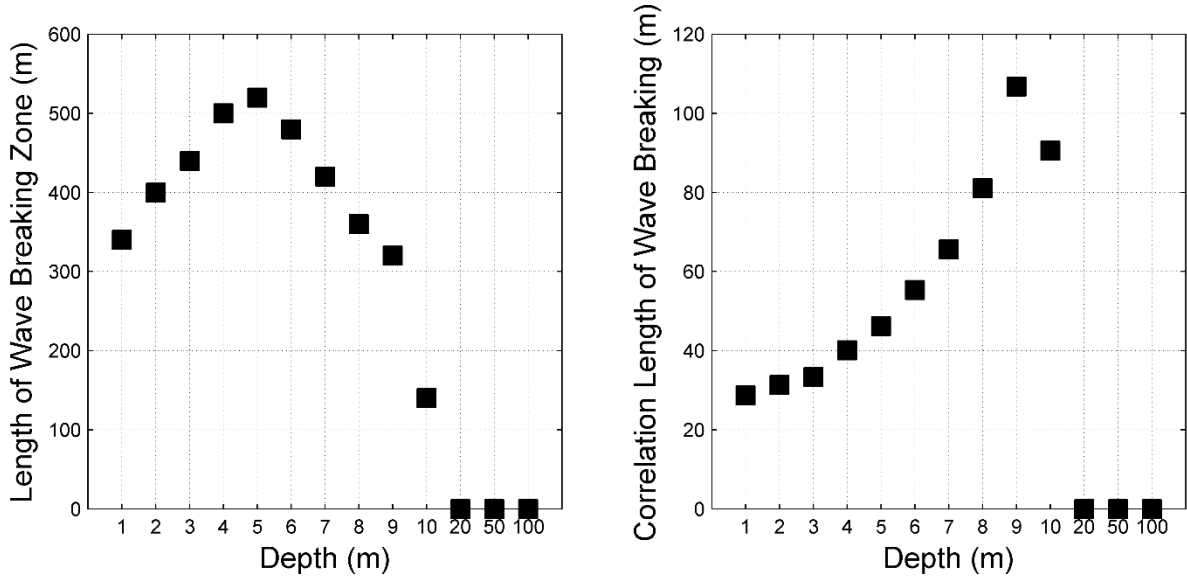
Figure 14. (Left) The length of the wave breaking zone as a function of the depth, and (Right) the estimated correlation length of breaking wave field as a function of the depth.

Not surprisingly, the SWAN implementation data in Figure 14 confirm that the length of the breaking zone depends on the ratio of the wave height over the local depth, because all the simulations are forced with the same climatological wave data from the N. Atlantic (wave height, period and direction), and the only variable is the depth. So for these wave conditions in shallow areas ($d \leq 5\ m$), the value of the integrated covariance increases because SWAN sets the total energy to the level that it can propagate into the grid without discontinuities. For $d > 5\ m$, the covariance decreases with increasing depth. This is due to an increase in the percentage of non-breaking waves in deeper water, as confirmed by the disappearance of the breaking zone for $d > 20\ m$. In order to estimate the spatial correlation length, Eq. (3.6) is applied to the normalized second derivative. From Figure 14 (right), it is clear that the correlation depends on the breaking ratio in space. In shallow areas, the energy of the wave field is brought almost instantly to the acceptable level, and therefore, the correlation length is short, roughly 2 grid points. As the depth increases, the wave field breaks over longer distances; thus, the correlation length increases.

The second step for the determination of the covariance function in the breaking zone is the fitting of the appropriate function. As is obvious from the top left of Figure 12, in general the

covariance of the breaking wave field declines rapidly with distance. With this in mind, we will investigate several formulations based on the general exponential function:

$$C_r^{Br} = \alpha_{Br} \exp(-\beta_{Br} r^{d_{Br}}) + c_{Br} \tag{3.11}$$

where C_r^{Br} is the covariance in the *x*-direction for the breaking wave field as a function of the distance *r* and α_{Br}, β_{Br}, c_{Br} and d_{Br} are fitted parameters that depend on the depth.

The most accurate function of the normalized covariance for each depth is when all the four parameters are fitted separately for each depth. In this case, RMSE values are on the order of 0.003 and SSE values are on the order of 10^{-4} (complete results not shown). In order to optimize a generalized empirical covariance function, the two definitions are proposed:

$$c_{Br} \equiv C_d \tag{3.12}$$

where C_d is the covariance function for each specific depth and

$$d_{Br} \equiv 1 \tag{3.13}$$

Under these two assumptions, the parameters α_{Br}, β_{Br} are estimated by fitting the data for each depth. The trend is modeled by simple exponential functions (Figure 15). Both parameters are equal to zero for depths exceeding $20m$. The goodness-of-fit statistics for the two parameters confirm that the parameters depend on the depth and that the use of exponential functions to represent them is reasonable.
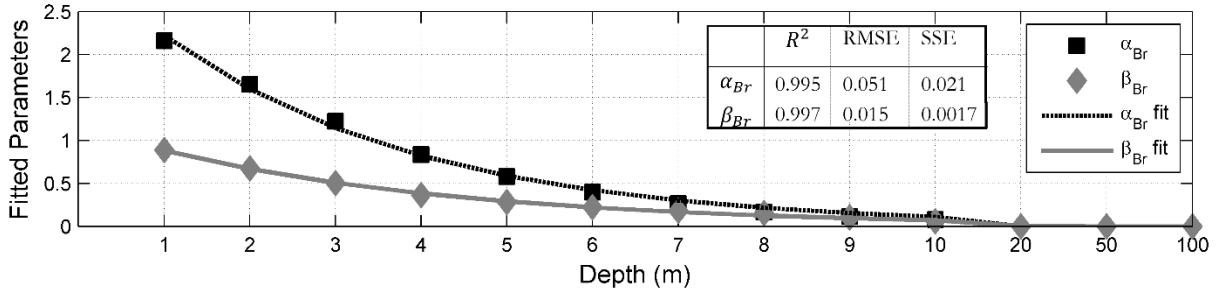
Figure 15. The calculated values of α_{Br} and β_{Br} of function (3.11) are plotted with squares and diamonds, respectively. The two lines correspond to the fitted models of the two parameters, and the different statistical measures of goodness-of-fit are tabulated.

Hence, the general function (3.11) is parameterized and the covariance function of the breaking wave-field induced by the local depth, *d,* is given by

$$C_r^{Br} = \alpha_{Br} exp(-\beta_{Br} r_x) + C_d$$

$$\alpha_{Br} = 3.099 \exp(-0.332d) \tag{3.14}$$

$$\beta_{Br} = 1.164 exp(-0.276d)$$

Equation (3.14) is validated against the original data of Figure 12a. The scatter plot of Figure 16 confirms the accuracy of the covariance function in comparison with the calculated covariance values. The slope of the linear best-fit line (gray) is 0.96, and its bias is -0.001.
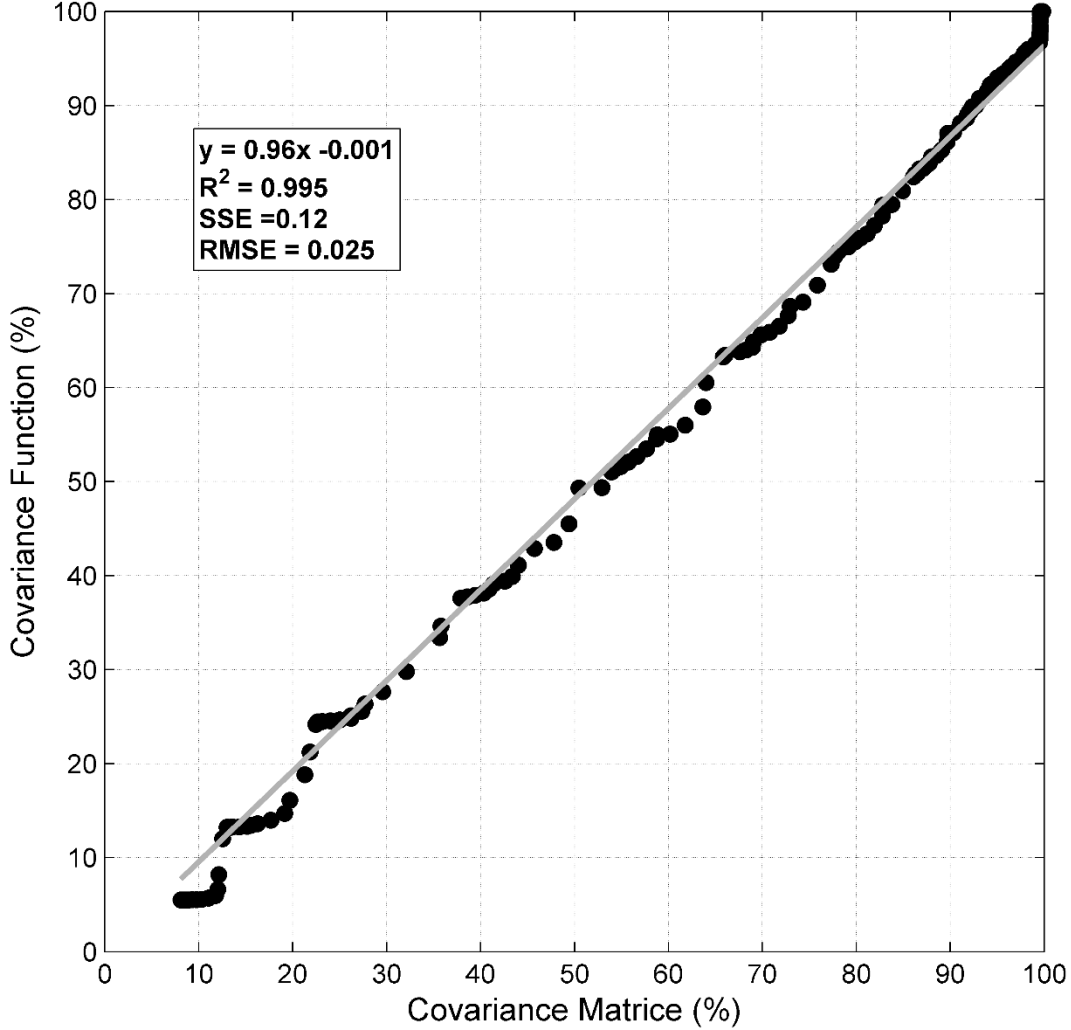
Figure 16. Scatter plot of the calculated normalized covariance versus the normalized covariance from the function (3.14).

The three measures of goodness-of-fit show that the two datasets are nearly identical. For values of relative covariance less than 20%, the covariance function tends to underestimate the simulation covariance. The main reason for this is the assumption (3.13), that the exponent is defined equal to 1 and therefore the value of C_r^{Br} tends faster to C_d. For values of relative covariance higher than 20%, the covariance function predicts the simulation covariance with greater accuracy. The remaining errors are considered acceptable in exchange for the simplicity of function (3.11), which has only two empirical parameters and minimal calculation cost.

3.2.4 Correlation Lengths of Covariance Function in Physical Space

In the direction parallel to the depth contours (i.e., the y-direction), the covariance of the wave energy is almost invariant for nearshore spatial scales $O(10km)$. Therefore, correlation length L_y is defined as equal to the size of the grid. The correlation length in the x-direction, L_x, is calculated indirectly via the correlation length in the z-direction (depth), L_d. As the covariance function in depth, C_d, is known, the correlation length is calculated by integrating its derivative in depth from $d = 0m$ to the depth at which $C_d = 1$, which in this case is $d = 20m$.

The derivative of (3.10) is calculated as

$$C_d' \equiv C_z' = 0.832408 exp(-0.375235\,z - 2.21837\,exp(-0.375235\,z)) \qquad (3.15)$$

and L_z, the correlation length in depth is estimated equal to $0.90m$. For an area with a known bathymetry, and consequently a known bottom slope, L_x is calculated using the bottom slope in the direction of maximum gradient for a specific location, $r(x_i, y_j)$. Thus, the spatially varying correlation length as function of the bottom slope in x-direction is given by:

$$L_x^r = 0.9/\beta_x \qquad (3.16)$$

3.3 Error Covariance Function in Spectral Space

Little has been published about the error covariance of the wave field in spectral space. Only Breivik et al. (1998) import the frequency indirectly by varying the spatial correlation length as a function of frequency under the dominant assumption of isotropy. However, they do not consider the covariance a function of direction. In this section, the covariance of the spectral wave action is examined and parameterized as a function of both frequency and direction.

The transfer of wind energy to the waves occurs mostly near the peak of the spectrum and at mid-range frequencies. The gained energy, through wave-wave interactions is (i) absorbed at lower frequencies (resulting in a shift of the peak frequency and generating infra-gravity waves) and at higher frequencies (possibly creating a secondary peak) and (ii) dissipated by white-capping and surf-breaking at all frequencies and by bottom friction at the low and mid-range frequencies. The objective is to determine statistically the relative importance of each of these mechanisms.

Let N and N' be two sets of action density spectra with the same length, M_t, sampled at two different locations from M_t different wave fields, propagating in the same coastal area. For simplicity, the action density spectra have been integrated over directions, and the values of N and N' are determined at two different locations of the same shoaling wave field. The ensemble is composed of time series of shoaling waves with varying H_s, and peak frequency, f_p (Figure 17). The transparent spectra correspond to ensemble N and the shaded spectra to correspond to N'.
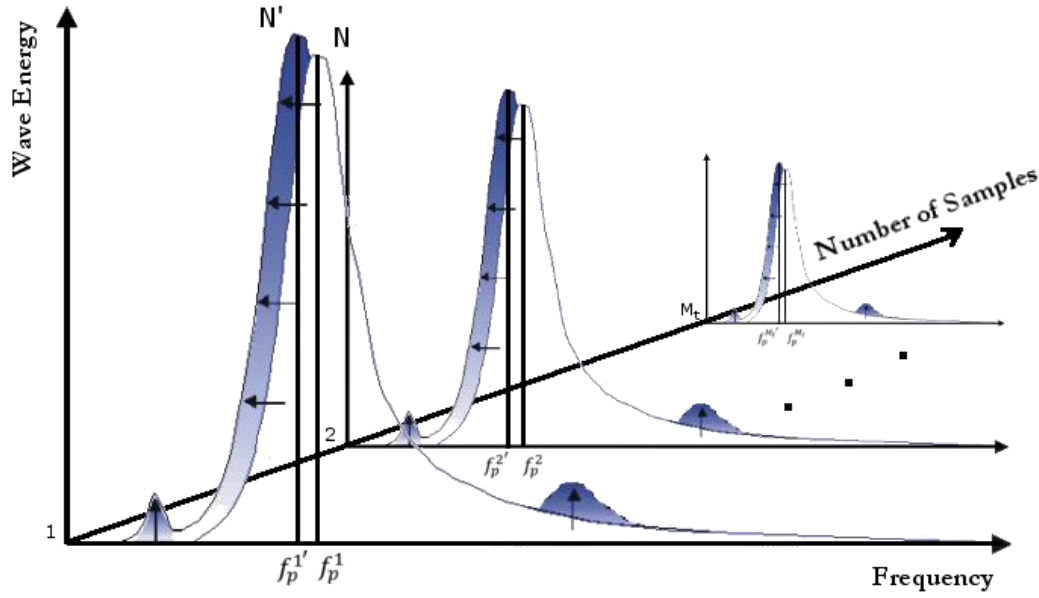
Figure 17. Schematic representation of ensemble spectra from two datasets N (transparent) and N' (shaded), showing input frequency spectra for the calculation of the shoaling wave field; modified from Holthuijsen (2007).

The error covariance spectra in frequency for two points are calculated between their two ensembles of two-dimensional spectra assuming that the spectra of ensemble N, open sea, have one f_p, the corresponding spectra of the shallower ensemble N' have up to three peaks (with two of them significantly smaller than the primary peak). f_p', for each N' spectrum is shifted towards zero, and its energy is higher than that at the peak frequency of the corresponding deep sea spectrum. Hence, the expected behavior of the frequency covariance spectra is high positive covariance at f_p' and negative covariance at the frequencies of the two smaller peaks.

In the next sections, we will use the data to guide the analysis and the development of the covariance functions. First, the general properties of covariance spectra as functions of frequency and direction are investigated using the SWAN dataset from Duck, NC (dataset "C" from Section 2.2.1). Then, expressions for covariance with frequency and covariance with direction are derived separately treating the two quantities as independent parameters, similar to the approach used for $C_{xx'}$ and $C_{yy'}$ in Section 0.

3.3.1   Covariance Spectra in Frequency and Direction

The covariance has been calculated between the wave spectrum at the central grid point of all grids and all the grid points of each grid: $C(N(f, \theta, 50,1), N(f', \theta', i_x, 1))$, where N is the action density at the grid node i_x. $i_x = 50$ is the central grid node, and each spectrum has 36 directions, θ, and 33 frequencies, f. Calculation of the covariance increased the data volume by approximately 33×36 times in comparison with the spectral output of SWAN for the same grid. Therefore, the challenge is the identification of the main characteristics of the covariance spectra in a computationally feasible way.

29

The first step in the analysis is the integration of the covariance spectra in frequency and in direction. For each spectral bin (f, θ) of N, a covariance spectrum for each grid point is estimated. The one-dimensional covariance spectra in frequency and direction with the maximum covariance are plotted for a range of depths in Figure 18. For these cases, the maximum covariance is used because the aim is to model the maximum error covariance for the assimilation system. The covariance for each frequency and direction depends on the depth, and the covariance of breaking waves is higher than the covariance of the propagating wave field. Both of these issues were addressed in Section 0 where a parameterized form for the function of the integrated covariance was presented. So, here, the covariance function will be developed based on covariance values normalized by the maximum covariance.

As may also be seen from Figure 18, the covariance for a given frequency and direction is almost constant along the grid (excluding grid nodes affected by the boundary). The covariance density is concentrated in a few specific frequencies (near 0.08 Hz) and directions (roughly normal to the boundary). The maximum directional spread is approximately 90°, and the difference between the maximum covariance density and the covariance at other frequencies and directions can be significant, up to four orders of magnitude. These observations are valid for the covariance of all the $N(f, \theta)$, not just for the peak frequency and direction of the covariance spectra. Based on these results, the dimensions of the covariance spectra will be further reduced by calculating their average in the x-dimension, excluding the boundary effect zone.
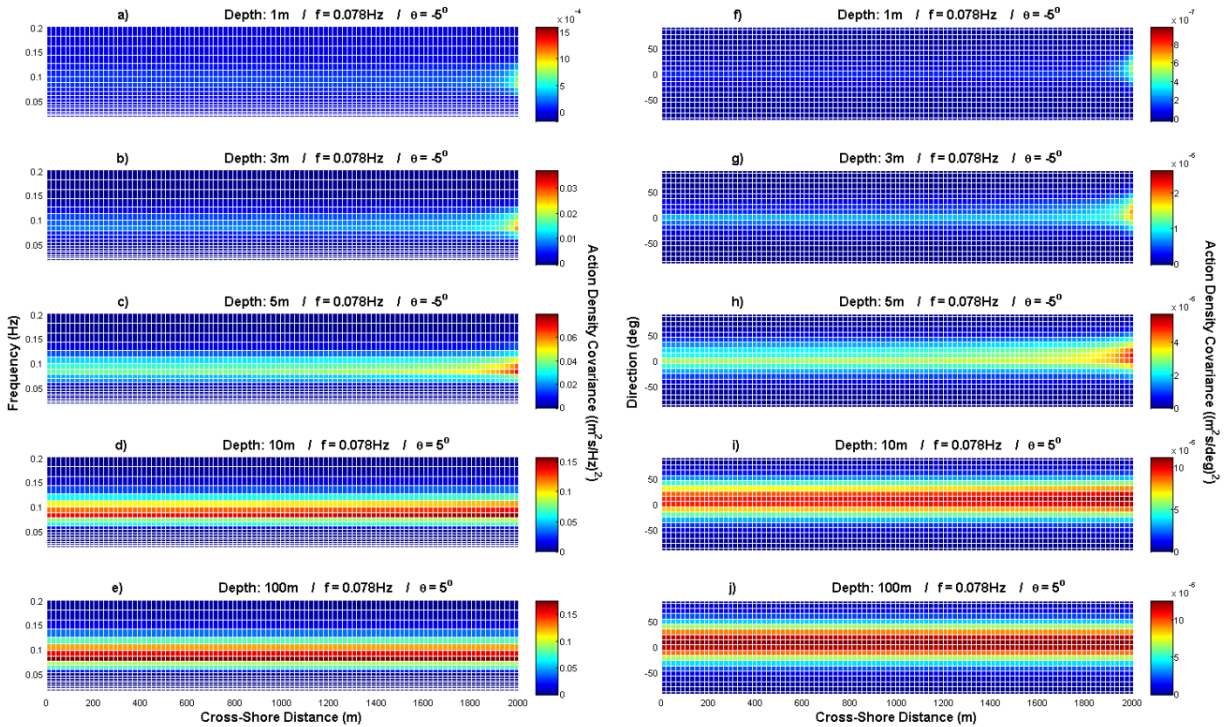
Figure 18. Spatial distribution of the maximum covariance as a function of frequency (left) and direction (right) for different depths (1m, 3m, 5m, 10m, 100m).

Prior to averaging in the x-dimension, the relative difference in physical space is calculated and averaged for all the combinations of f and θ, in order to quantify the spatial variation of the

covariance for each f' and θ' and for all the grids (Figure 19). For grids with depth less than $10m$, the maximum difference is observed at the frequency, $f' \approx 0.08Hz$ and direction, $\theta' \approx 0°$ of the covariance peak. The order of magnitude of the maximum values of the covariance difference for all the depths is four to six times smaller than the actual values of the covariance, which may be seen by comparing the scales of Figure 18 and Figure 19. The difference of the covariance for the rest of the spectral bins is practically zero (dark blue in Figure 19). For grids with depth more than $10m$, the difference at the covariance peak is negative for the integrated covariance over frequency. There are two extremes for the integrated covariance over direction with a positive value at $f' \approx 0.15Hz$ and a negative value at $f' \approx 0.08Hz$ (Figure 19e and j). This is an expected result of the energy transfer mechanisms in SWAN. In both cases, the order of magnitude of the difference is approximately five times less than the order of magnitude of the covariance. Figure 19 suggests that the covariance spectra may be spatially averaged without eliminating significant information. Moreover, it is also apparent that specific frequencies and directions could be eliminated from the rest of the analysis, in order to reduce the data volume further.
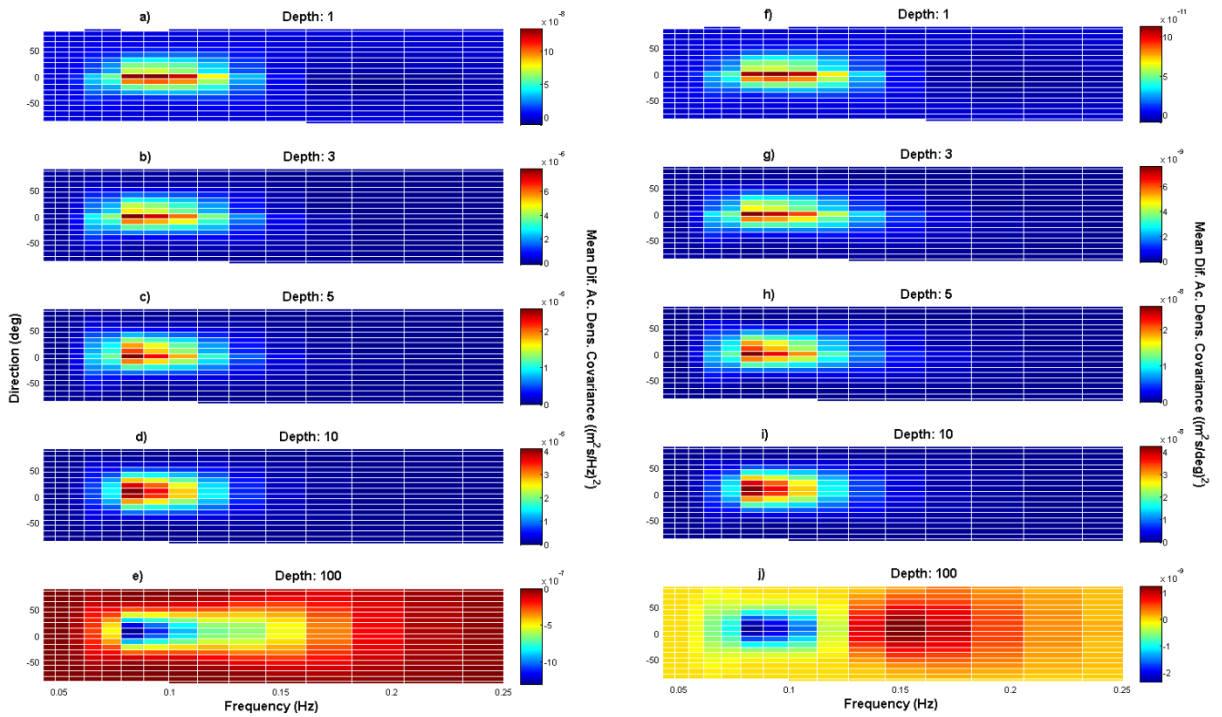
Figure 19. Mean spatial difference of the covariance spectra for each frequency and direction (f' and θ') for spectra integrated over frequency, f, (left) and direction, θ, (right).

The next step in the process is the identification of the frequencies and directions that can be eliminated. For this purpose the spatially averaged covariance spectra are used, and covariances in frequency and in direction are both examined in the same manner.

The covariance spectra, $CS(f', \theta')$ are examined for different values of frequency, f, and direction, θ, in Figure 21 and Figure 22, respectively. As discussed before, the covariance has been calculated at only the points where the action density was higher than the resolution of the

31

computational system, resulting in directions for which the covariance has not been calculated. Of the 33 frequencies of the original data, only the 16 frequencies, $0.0426Hz \leq f \leq 0.261Hz$, have more than 0.1% of the maximum covariance for the specific conditions of the simulations. If the threshold for the covariance density is increased to 5%, the number of frequencies with significant covariance density is reduced to 10, $0.0543Hz \leq f \leq 0.161Hz$. From Figure 21, it is clear that the direction of the peak covariance is $\theta' = 0$, normal to the coast, for all values of the direction, θ. In all the plots of the two figures, 5° is the exact direction corresponding to the peak of the covariance. This is an artifact of having 36 directional bins of 10° in SWAN. In cases where an even number of directions is specified (as done for the present analysis), this results in directional bins at $+d\theta/2$ and $-d\theta/2$ but not at zero. The maximum directional spread is 90°, as expected. The exact directional spread of the covariance spectra is discussed in section 3.3.2, where the error covariance function in direction is determined. In both Figure 20 and Figure 21, it is clear that the distribution of the covariance in frequency is positively skewed and it is symmetric in direction. These two properties are fundamental for the development of the covariance functions.
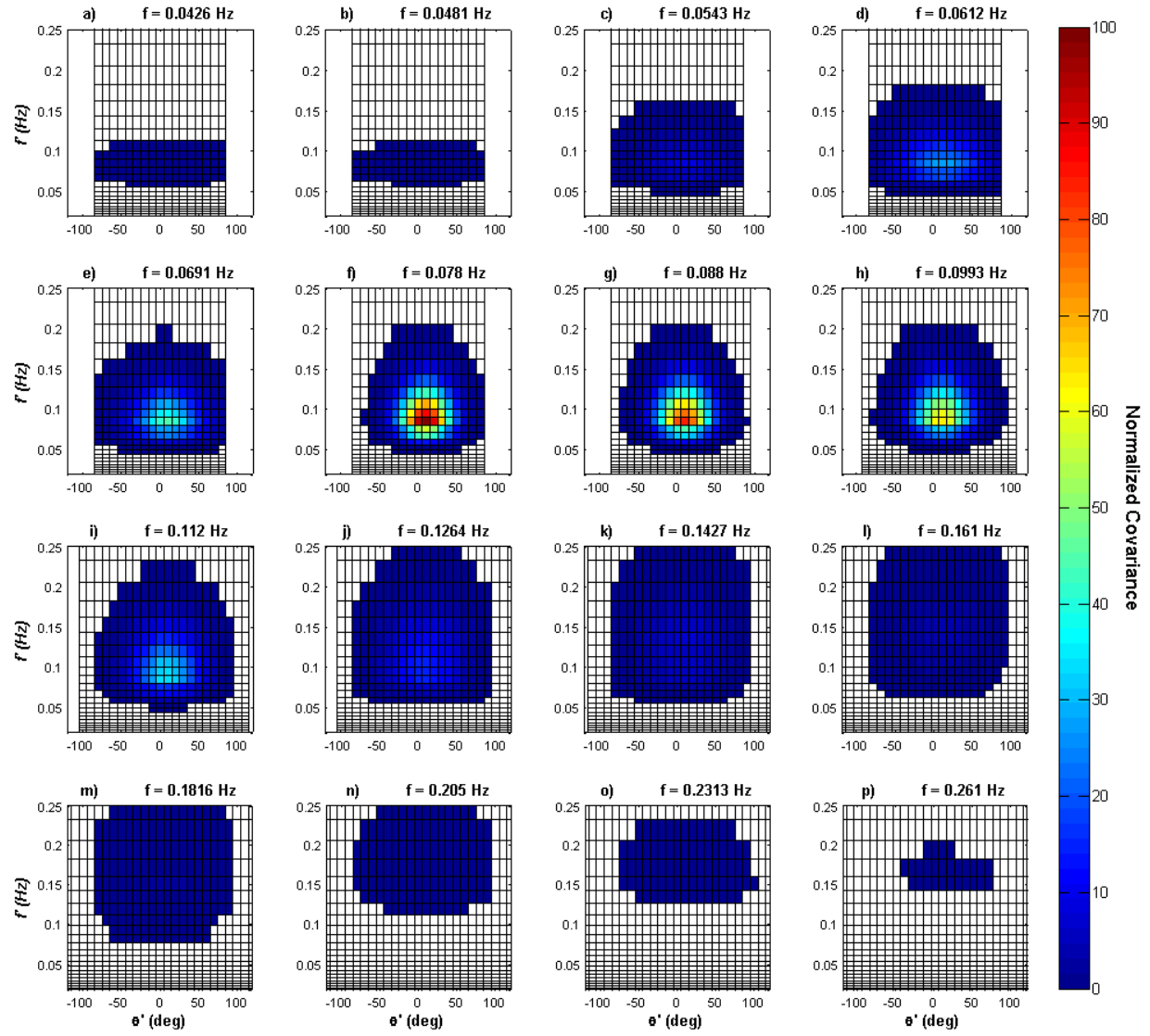
Figure 20. The normalized covariance spectra in frequency and direction for each frequency of the covariance ensemble at the point (0*m*, 1010*m*) of the grid, depth 10*m*. The angle 0° corresponds to the direction normal to the wave field propagation.
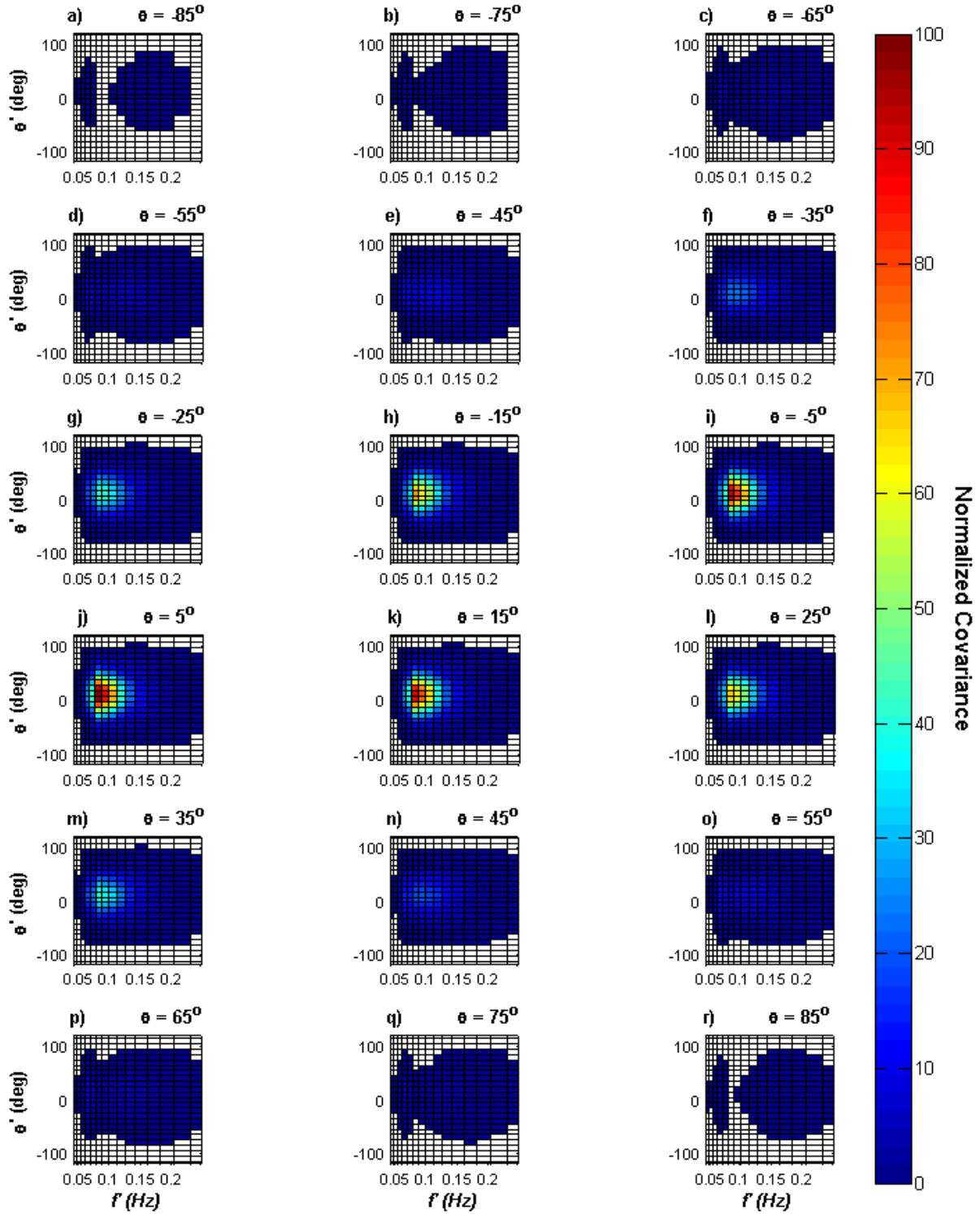
Figure 21. The normalized integrated covariance for all the directions θ for which, the covariance has been calculated as a function of each direction θ' and frequency f'. The depth is $10m$, and the direction $5°$ is the normal to the boundary. The missing directions are excluded from the covariance calculation.
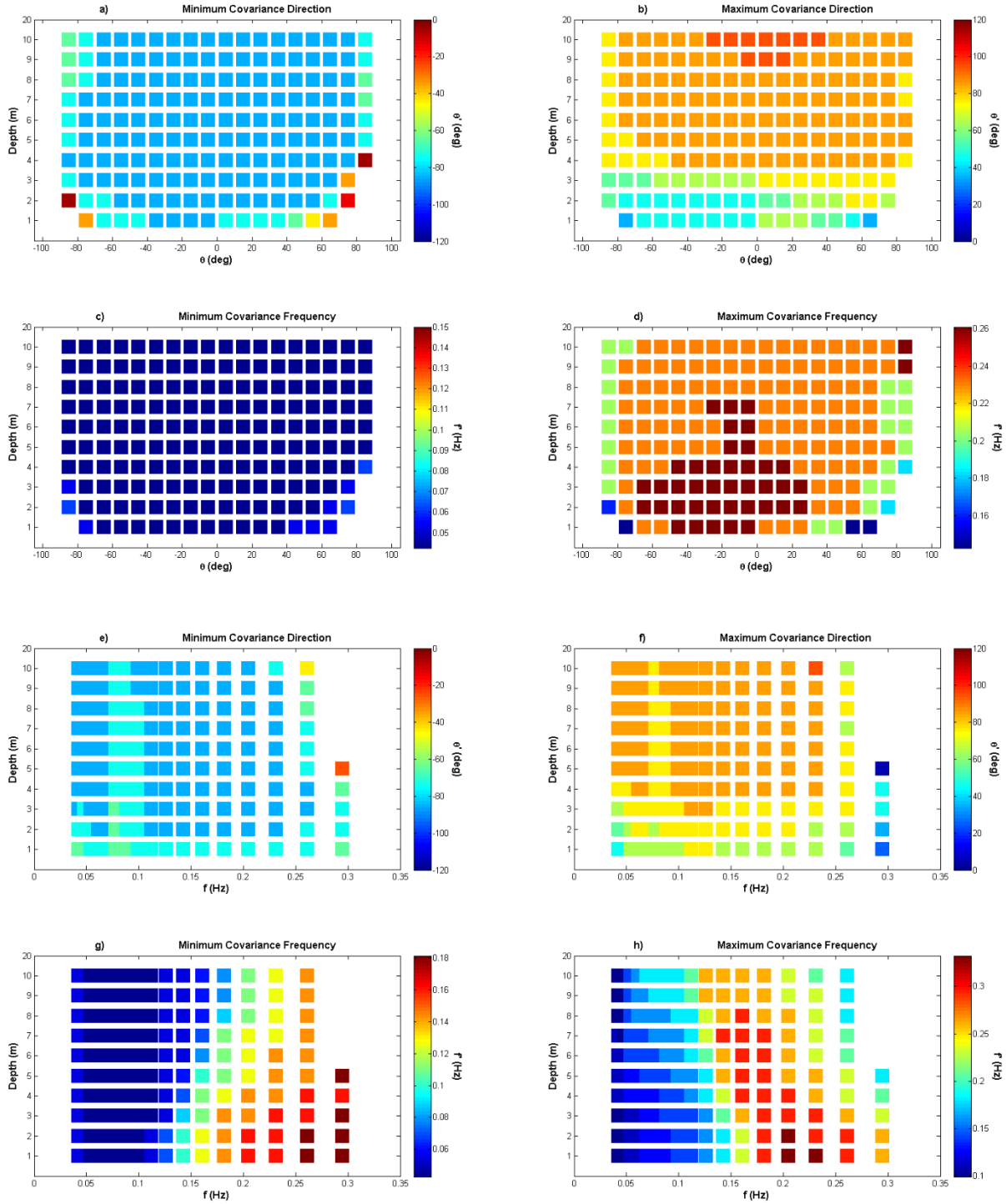
Figure 22. Summaries of minimum (left) and maximum (right) frequencies and directions per depth with significant spectral density.

Figure 22 summarizes basic information about the frequency width and directional spread based on minimum and maximum values as functions of f and θ without applying any threshold. The

figure includes the information from all covariance spectra for each depth. Therefore, based on its content, the results from the observations at $10m$ depth, as shown in Figure 20 and Figure 21, are generalized. The main direction is the normal to the coast, and the included frequency range is $0.04Hz \leq f \leq 0.25Hz$. There is a dependency on the actual depth for both variables, but for the development of the covariance functions, the covariance spectra are examined according to the covariance density as a function of frequency width and directional spread, instead of the extreme values of frequency and direction.

3.3.2   Development of Error Covariance Function

In the next sections, the covariance in spectral domain is examined independently in frequency and direction:

$$C_{NN}(f, f', \theta, \theta') = C_{ff'}C_{\theta\theta'} \qquad (3.17)$$

Here $C_{ff'}$ is the covariance between the frequencies f and f', and $C_{\theta\theta'}$ is the covariance between the directions θ and θ'. In addition, the covariance, $C_{\theta f'\theta'}$, as a function of f, θ and θ' will be presented.

3.3.2.1   Error Covariance Function in Frequency $C_{ff'}$

Analysis of the covariance in frequency domain showed that covariance as a function of the frequency is positively skewed. The directionality of the covariance spectra, e.g. Figure 20, has been neglected by integration over direction. The top panel in Figure 23 is a characteristic example which depicts the frequency space distribution of the covariance for depth of $10m$, calculated for the spectral bin corresponding to $f = 0.0612 \ Hz$, $\theta = 185°$. It shows that for the whole length of the grid, the peak is at the same frequency and the covariance is almost constant in space. This confirms the conclusion based on the integrated covariance that in the x-direction, the covariance is almost constant when the bathymetry is homogeneous. In order to estimate the difference in space, the average of mean relative differences in cross-shore direction for each spectral bin have been calculated for all the depths (Figure 24). The difference has been calculated for the grid points that are not impacted by the boundary effect, and the mean relative difference is 0.001% with the bars showing the standard deviation. Even though the mean for all the depths is close to zero, the relative standard deviations have high values due to the high frequencies (higher than 0.2 Hz, the blue areas in Figure 23) where the values of the covariance are extremely small which makes the ratios relatively high and consequently high values for the standard deviation. Nevertheless, Figure 24 proves that there is not significant spatial variation of the covariance spectra. Therefore, the mean covariance in space for each spectral bin is used for the covariance function. Hence, for the next steps the covariance spectra are averaged in space, e.g. Figure 23 bottom, and the averaged spectra are used for the next steps of the analysis. The integrated covariance spectrum confirm that the values of the covariance for frequencies higher than 0.2 Hz are significantly smaller (up to orders of magnitude) in comparison with the maximum values.
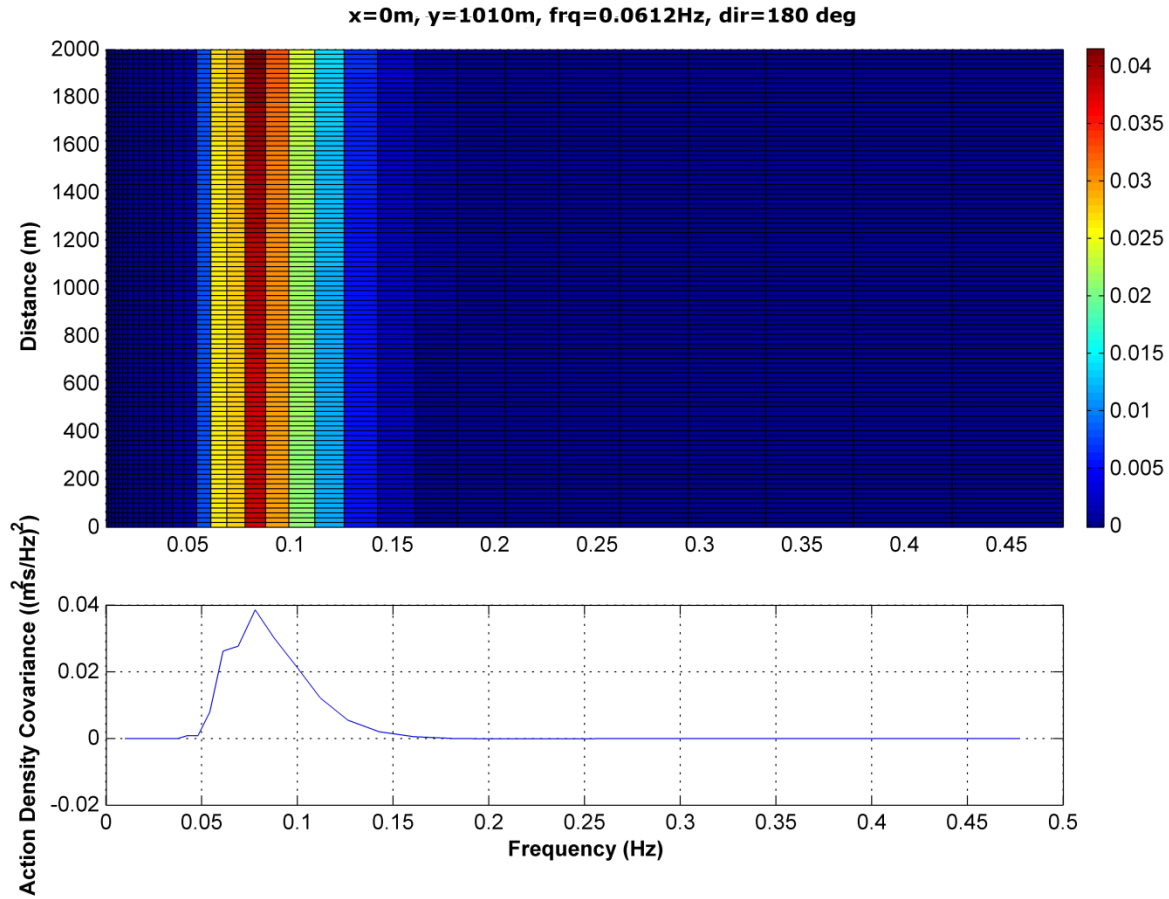
Figure 23. Top: Frequency – Space Distribution of the maximum Covariance Action Density calculated for 10 m depth; it corresponds to the spectral bin (0.0612 *Hz*, normal to the shore) of grid point 0 m, 1010 m). Bottom: The same data for covariance frequency spectrum integrated over space.
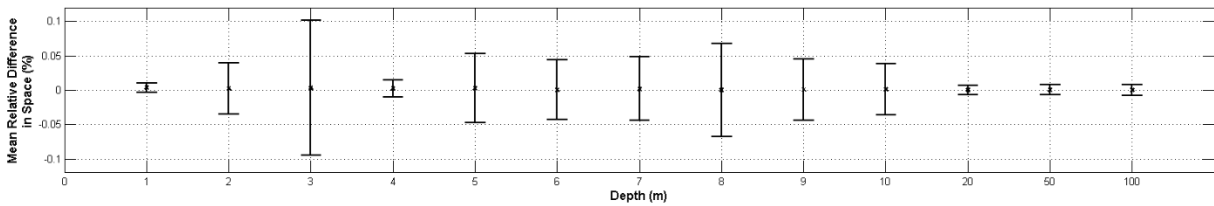
Figure 24. The integrated mean and standard deviation of the relative difference in space for all the frequencies and depths.

In summary, the covariance spectra for all the frequencies have been integrated over direction, and the frequency covariance spectra as a function of the depth are shown in the left side of Figure 25, which confirms the strong depth dependence and that the peak frequency of the covariance is almost constant for all the depths.
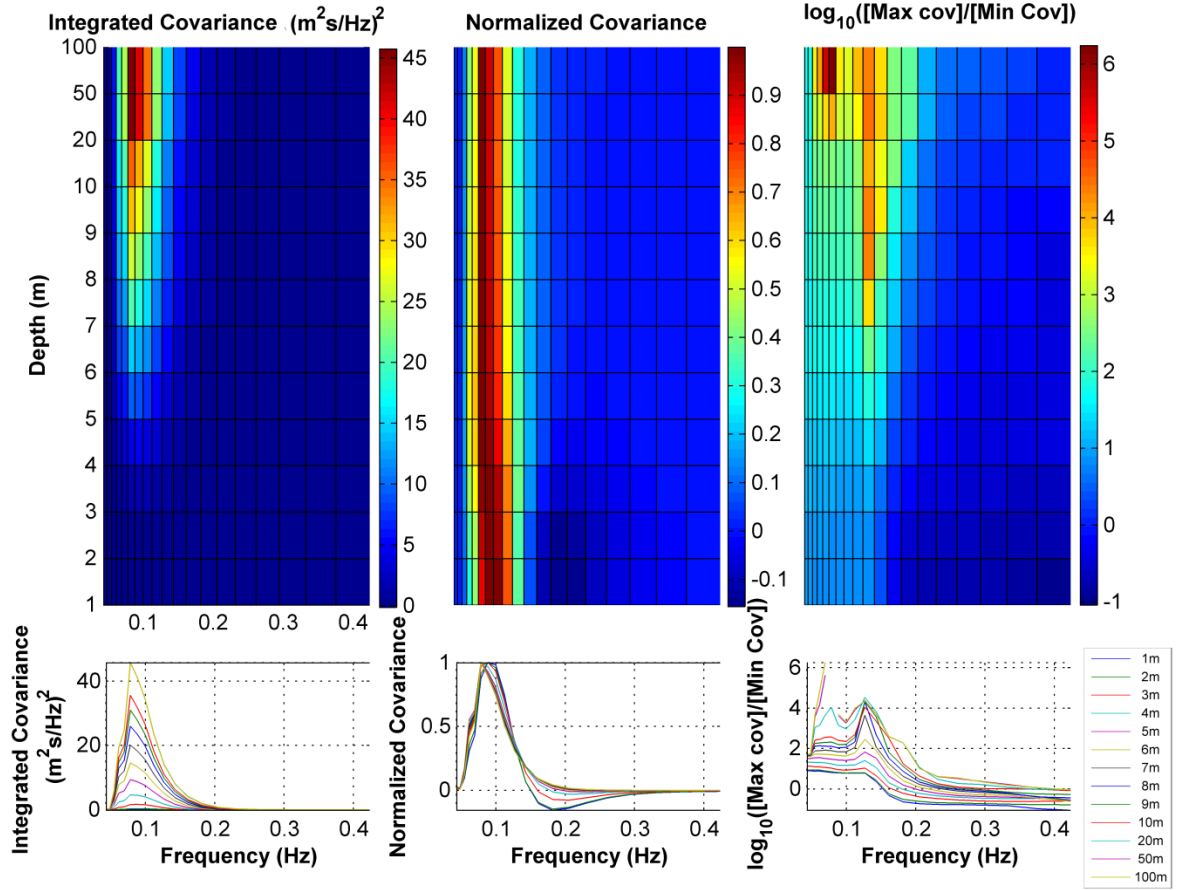
Figure 25. Left: Integrated over direction, covariance for all the spectral bins of point ($0m$, $1010m$) as a function of the frequency and depth. Middle: Normalized, integrated covariance with the maximum of each depth, spectra. Right: The ratio of extreme (max/min) covariance in logarithmic scale.

Using the covariance spectra at the peak frequency of the integrated covariance spectra (lower left plot of Figure 25), the function that fits the maximum value of the normalized integrated covariance is:

$$\max\left(\sum_f C_{ff'}\right) = \begin{cases} 1, & d \geq 10m \\ 0.0115d^{13/7}, & 0m \leq d \leq 10m \end{cases} \qquad (3.18)$$

where d is the depth. Equation (3.18) is produced by fitting the power law to the maxima integrated covariance, yielding $R^2 = 0.9902$, $RMSE = 0.0050$ and $SSE = 0.0002$.

Due to the absolute difference between the covariance at shallow and deep grid points, the frequency covariance spectra have been normalized by their maximum value and are plotted as a function of the depth (Figure 25, middle). The normalized covariance spectra for all the depths have similar shape in frequency domain. The peak frequency and the frequency width are

similar, and their distributions are positively skewed towards high frequencies.Additionally, the covariance tends to zero. More precisely, the peak frequency is approximately at 0.09 Hz, and most of the covariance density is concentrated in less than 10 frequency bins. The only exception is for $d \leq 3m$. The covariance frequency spectra become broader. The peak frequency is shifted towards higher frequency, and for frequencies higher than 0.161 Hz, the covariance becomes negative. After the frequency 0.3 Hz, it tends to zero. In order to determine the maximum variation of the covariance, the ratio of the maximum value of covariance over the minimum for each frequency and depth was calculated (Figure 25 right). The scale is logarithmic, and we see that for frequencies smaller than 0.25 Hz, the extrema are up to 3 orders of magnitude smaller than the values at the peak frequency. Hence, the rest of the analysis will focus on frequencies between 0.0426 Hz and 0.2945 Hz.

Based on the normalized covariance spectra (Figure 25, middle) the following function was fitted to the data for each depth:

$$C_f = a_{f1}e^{-(f-b_{f1}/c_{f1})^2} + a_{f2}e^{-(f-b_{f2}/c_{f2})^2} \tag{3.19}$$

where $a_{f1}, b_{f1}, c_{f1}, a_{f2}, b_{f2}, c_{f2}$ are the fitted parameters. Equation (3.19) is the sum of two Gaussian distributions, and it is used in order to model the positive skewness of the covariance as a function of the frequency. The fitted parameters for all the depths are given in Table 2.

Table 2. Fitting parameters for the normalized covariance values as a function of frequency C_f.

	Depth (m)												
	1	2	3	4	5	6	7	8	9	10	20	50	100
a_{f1}	1.05	0.88	0.64	0.66	0.70	0.72	0.73	0.74	0.75	0.75	0.74	0.75	0.75
b_{f1}	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
c_{f1}	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
a_{f2}	-0.15	0.58	0.70	0.63	0.55	0.51	0.48	0.46	0.44	0.43	0.43	0.43	0.43
b_{f2}	0.06	0.11	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
c_{f2}	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.03
R^2	0.96	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
RMSE	0.08	0.07	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
SSE	0.09	0.08	0.04	0.06	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04

The statistics in Table 2 imply that Eq. (3.19) is appropriate to model the average covariance over all the frequencies and almost constant for all $d \geq 3m$. Similarly, the fitted parameters are almost constant for the same depths. Therefore, the parameters for $d \geq 3m$ are averaged in order to create one covariance function independent of the depth, which gives

$$C_f = (0.72 \pm 0.04)e^{-\left(\frac{f-(0.08\pm0.00)}{(0.02\pm0.00)}\right)^2} + (0.50 \pm 0.09)e^{-\left(\frac{f-(0.11\pm0.003)}{(0.03\pm0.003)}\right)^2} \tag{3.20}$$
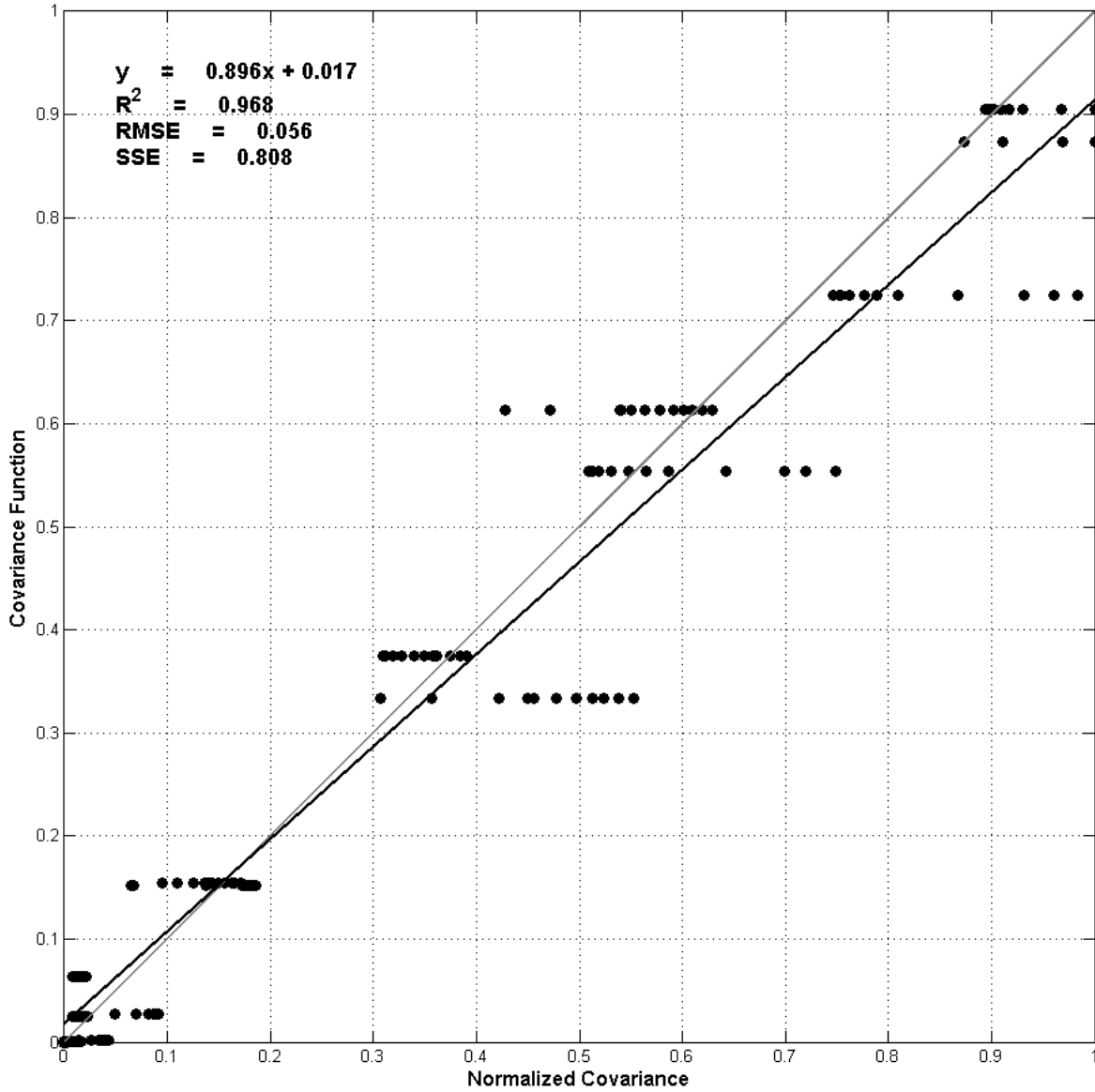
Figure 26. Scatter plot of the average normalized covariance for each of the depths versus the mean covariance function C_f.

As shown in Figure 26, the oversimplified function (3.20) of the covariance is relatively accurate. The goodness-of-fit statistics are also shown in the plot. Apparently, as long as the function does not depend on the depth, there are multiple values of the normalized covariance for each value of the covariance function. This approach has extremely limited computational cost, and it is directly applicable to any wave assimilation system as long as a directional spreading function is assumed.

Beyond the simplified covariance function given by Eq. (3.20), the objective is to determine a covariance function depending on both frequencies, f and f'. Recently, a covariance function,

$C_{ff'}$, based on the statistically normalized, covariance spectra, was introduced (Flampouris et al., 2014):

$$C_{ff'} = A_1 e^{-\left(\frac{f'-B_1}{C_1}\right)^2} + A_2 e^{-\left(\frac{f'-B_2}{C_2}\right)^2} \tag{3.21}$$

where A, B, C are the fitted parameters for each f. In this case, all the frequencies were used and the covariance function for each depth was created.
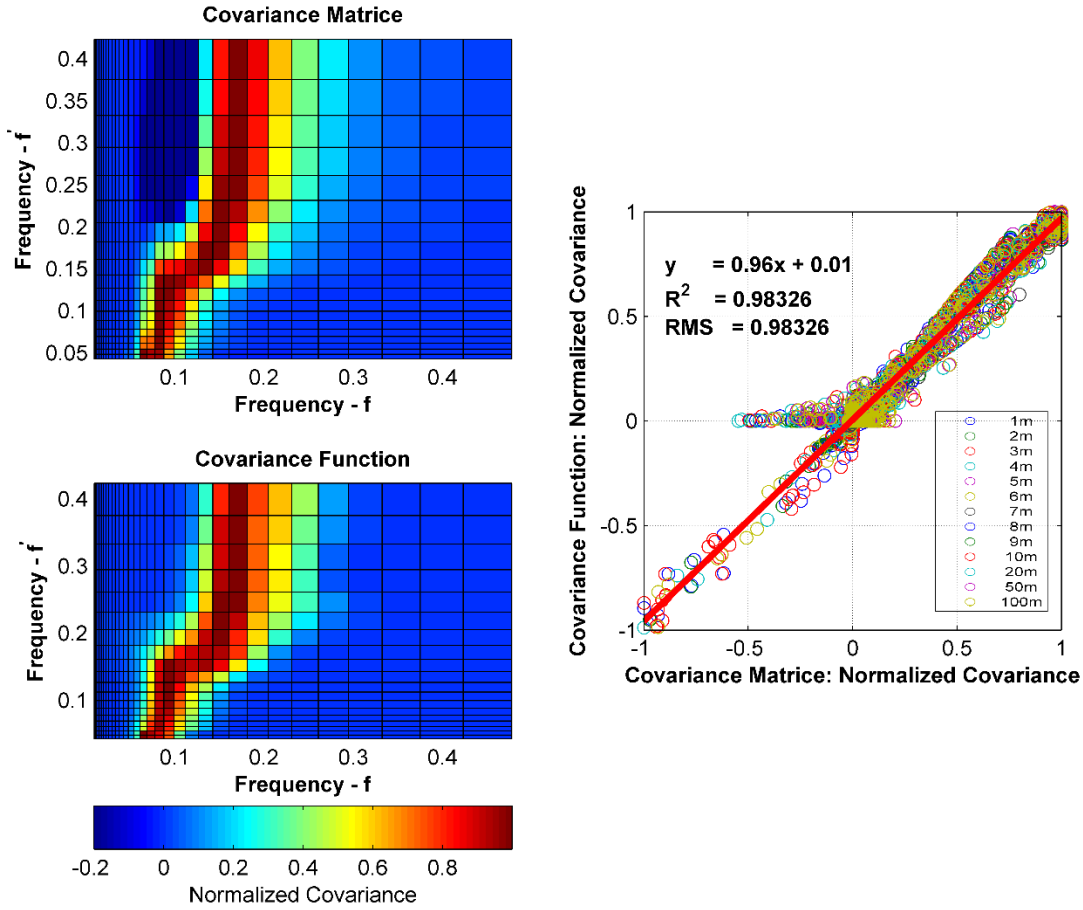
Figure 27. Left-Top: The calculated covariance for all the combinations of f and f'. Left-Bottom: The results of the covariance function, based on (3.21) for 10m depth. Right: The scatter plot between the calculated normalized covariance from the data versus the estimated covariance from the covariance functions.

This approach models the covariance accurately as shown in Figure 27. The main limitation of this approach is the partial modelling of the negative values of the covariance mainly between the frequencies of energy transfer. From a technical point of view, the limitation is due to the large number of fitted parameters.

The integrated covariance spectra normalized with the global maximum value of covariance at specific depths for all the combinations of f and f' are examined in Figure 28. On the global

scale of a shoaling wave field, the frequency covariance density varies significantly as a function of all the frequencies of the incoming wave spectra in a few and specific frequency bins. It is clear that at the shallower grid points, e.g. $1m$ depth, the covariance spectrum is more spread than at the grid points corresponding to deeper water. In Figure 28, the color scale is different for each plot. At the 1m depth, it is on $O(10^{-4})$, and at 10m depth, it is on the order of $O(10^{-0})$.
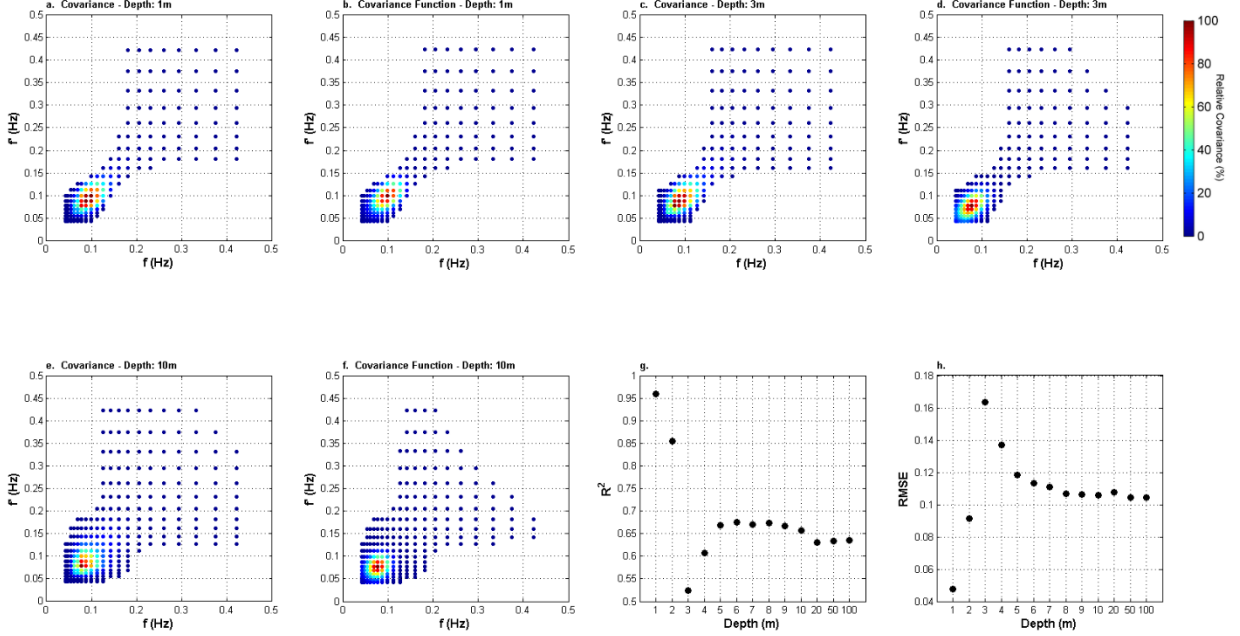
Figure 28. Covariance in frequency for all the depths. In all the graphs the x-axes correspond to the target spectrum frequency, f', and the y-axes correspond to the frequency of observed spectrum, f.

Based on the data of Figure 28, the covariance function, $C_{ff'}$ is determined. As discussed previously, there are no significant differences of the normalized covariance in frequency as a function of the depth, except in the peak frequencies. Hence, the median of covariance for each combination of frequencies (f, f') was calculated, and based on the median of covaieance, the following function was fitted:

$$C_{ff'} = a_f e^{-[b_f(f-f_p)^2 + 2c_f(f-f_p)(f'-f'_p) + d_f(f'-f'_p)^2]}$$ (3.22)

where a_f, b_f, c_f, d_f are the fitted parameters and f_p and f'_p the peak frequencies for f and f', respectively. The fitting procedure allows the function given by (3.22) to be simplified further by defining $a_f = 1$ and $b_f \approx d_f$, with differences less than 0.001%. Based on these two observations Eq. (3.22) is simplified further to depend on only two parameters. By following the same methodology as in the other parts of the covariance function, Eq. (3.22) was further simplified by parameterizing b_f and c_f as:

$$b_f = \begin{cases} 10^{3.152d^{0.01012}}, & d < 2m \\ 1490, & d \geq 2m \end{cases}$$ (3.23)

$$c_f = \begin{cases} -745, & d < 3m \\ -10^{3.276d^{-0.1203}}, & 3m \leq d \leq 10m \\ -305, & d \geq 10m \end{cases} \qquad (3.24)$$

The goodness of fit-statistics for the two parameters are provided in Table 3.

Table 3. The goodness-of-fit statistics of b_f and c_f for functions (3.23) and (3.24).

	R^2	RMSE
b_f	0.986	0.0030
c_f	0.998	0.0060

3.3.2.2   Error Covariance Function $C_{\theta\theta'}$

The covariance function in direction, $C_{\theta\theta'}$, is determined by following the same methodology as for the rest of the dimensions. The covariance in direction has been already examined qualitatively earlier where it was shown that $C_{\theta\theta'}$ is almost symmetric, with axis of symmetry the direction of the wave propagation. The effect of the frequency on the covariance spectra has been neglected by integrating over frequency. Figure 29 (top) depicts the direction-space distribution of the covariance for depth 10m, calculated for the spectral bin $f = 0.0426\ Hz$, $\theta = 185°$. The bottom plot of Figure 29 shows the spatially averaged covariance of the action density plotted for all directions. Since there are 36 directional bins with $d\theta = 10°$, the values in each bin are assumed to be at $d\theta/2$, which results in the maximum value of the averaged covariance at 5°. In order to estimate the variation in space, the average of mean relative differences in the cross-shore direction for each spectral bin have been calculated for all the depths (Figure 30). The mean relative difference is close to zero, and the bars show the standard deviation to also be small (<0.1%). The difference has been calculated for the grid points that are not impacted by the boundary effect. Nevertheless, Figure 30 proves that there is not significant spatial variation of the covariance spectra. Therefore, the mean covariance in space for each spectral bin is used for the covariance function. The integrated covariance spectrum confirmed that the values of the covariance for directions $\theta \leq |85°|$ are significantly compared to the maximum values.
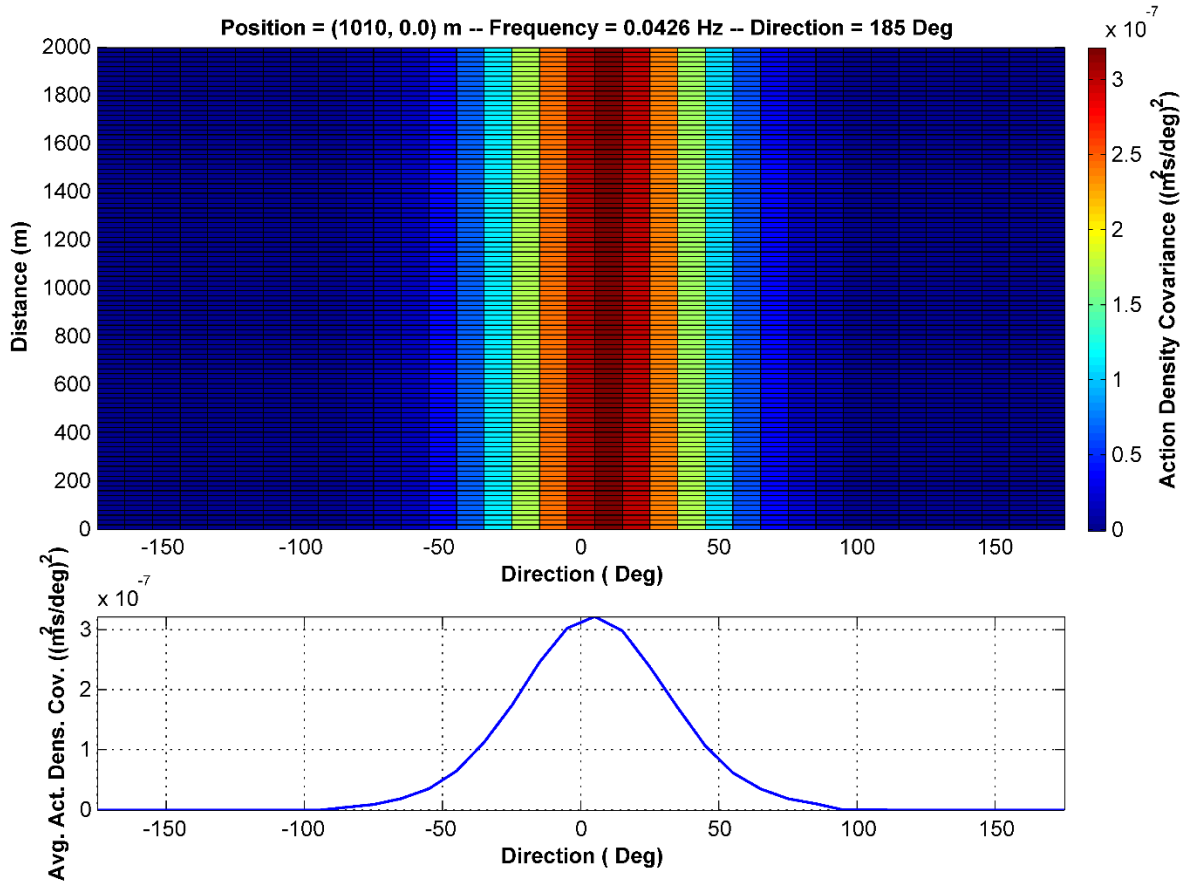
Figure 29. Top: Direction – Space Distribution of the maximum Covariance Action Density calculated for 10 m depth. It corresponds to the spectral bin (0.0426 *Hz*, 185 deg) of grid point (0 m, 1010 m). Bottom: Integrated over space, covariance/ direction spectrum.
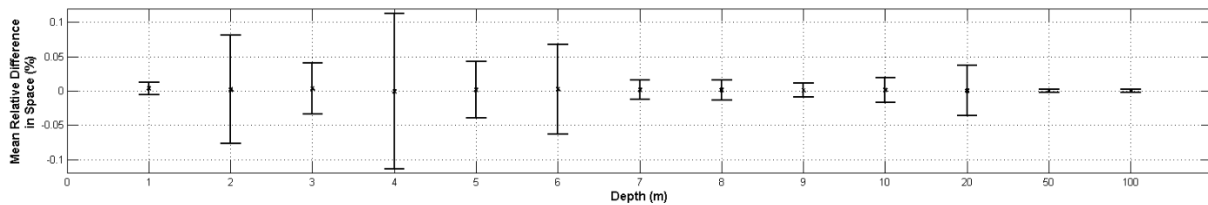
Figure 30. The integrated mean and standard deviation of the relative difference in space for all the directions and depths.

Next, based on the analysis illustrated in Figure 21, the directional spreading of the integrated covariance spectra was estimated for each depth. In cases where the maximum covariance was less than the 2.5% of the total covariance for the specific depth, the value has been neglected. Otherwise, as shown in previous section, the maximum directional spreading is 180°. Results of the current procedure are shown in Figure 31 and summarized for each depth in Table 4.
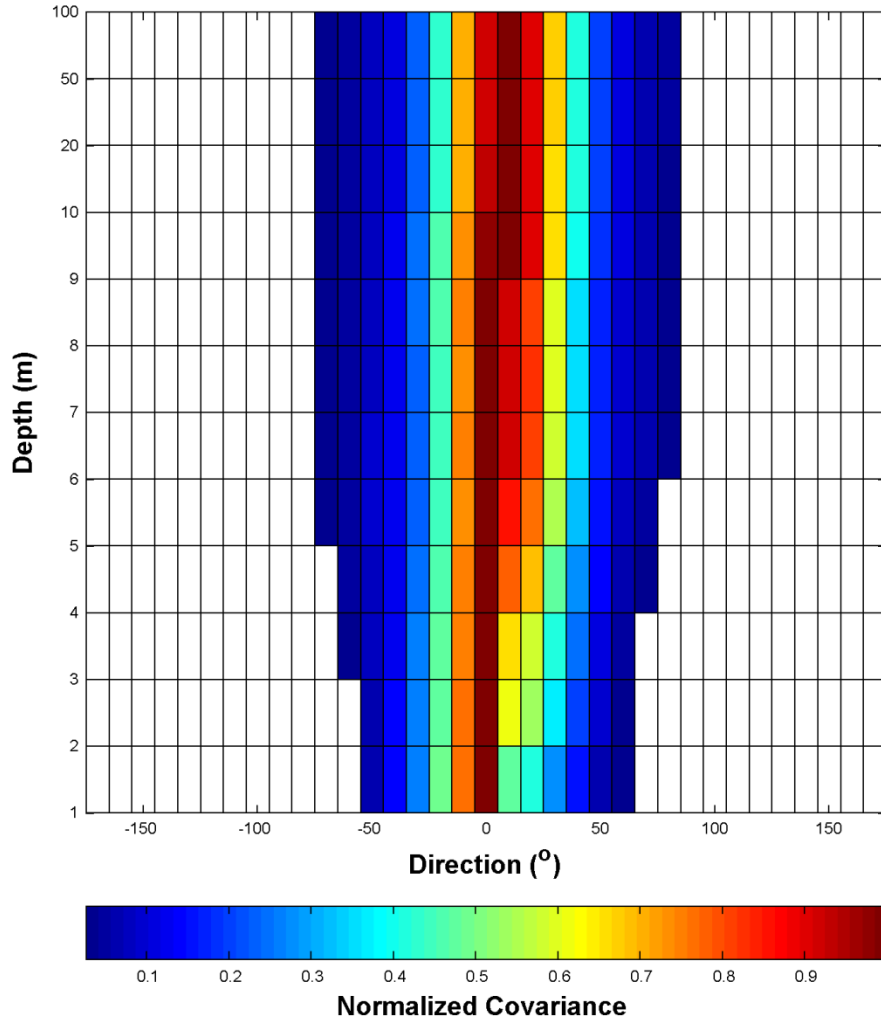
Figure 31. Maximum normalized covariance as a function of the direction and depth.

Table 4. Directional spread of maximum and average covariance as a function of the depth based on existence of data.

Depth (m)	< 3	3-5	> 5
Directional Spreading – Maximum (°)	55	65	75
Directional Spreading - Average (°)	55	65	75

From the covariance matrices in direction, it is obvious that the covariance density is concentrated on the peak directions for both θ and θ', especially in shallow water where the wave field directionality is dominated by the local bathymetry (Figure 32). From the same figure, it is clear that the $C_{\theta\theta'}$ have directionality, as well, in shallow water as evidenced in Figure 32 a-c, where the maximum values of the covariance are not symmetric around the peak directions. In contrast, the covariance matrices for depths greater than $9m$, Figure 32 i-m, are symmetric around the peak directions. Nevertheless, the maximum value of the covariance is at the peak

45

directions (θ, θ') and is reduced with the increase in direction. Based on the data shown in Figure 32, the covariance function, $C_{\theta\theta'}$ is determined using the median covariance spectra for each depth, integrated over frequency. To keep the properties of the covariance matrices, the following function was fitted:

$$C_{\theta\theta'} = e^{-\left[\left(\frac{\theta-\theta_p}{2\sigma_\theta}\right)^2 + \left(\frac{\theta'-\theta'_p}{2\sigma_{\theta'}}\right)^2\right]} \tag{3.25}$$

where $\theta_p, \sigma_\theta, \theta'_p, \sigma_{\theta'}$ are the fitted parameters.
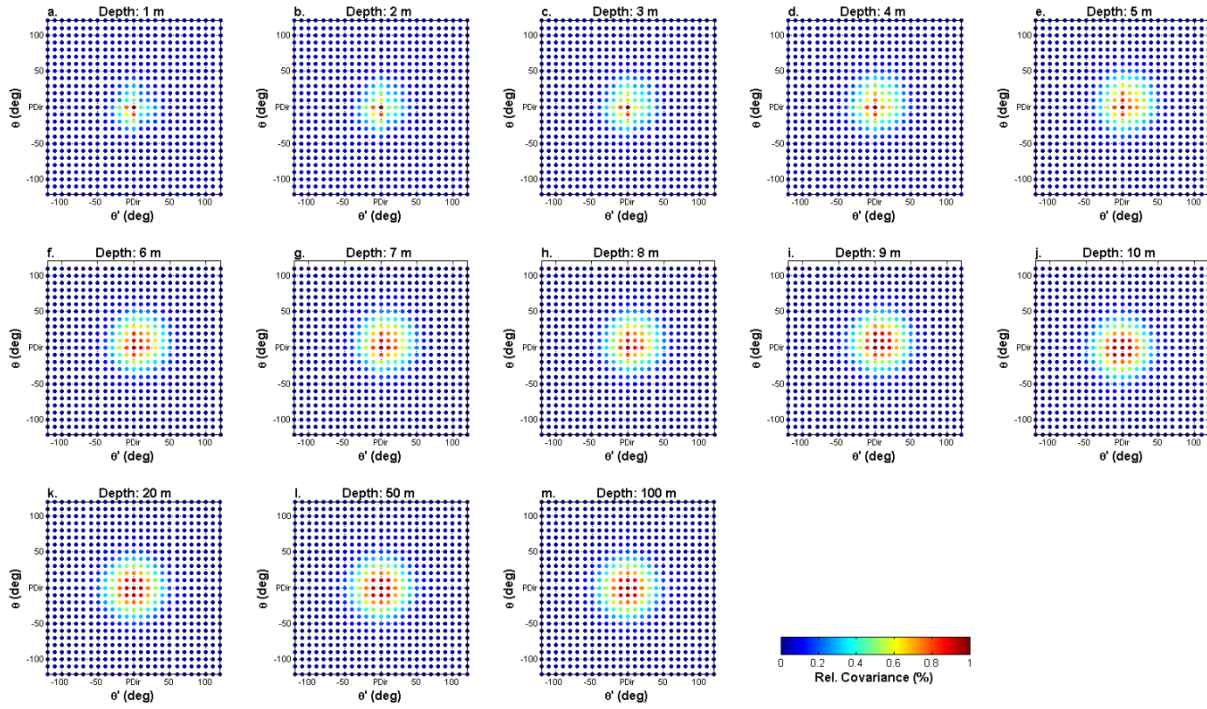
Figure 32. Covariance in direction for all the depths. The *x*-axes correspond to the target spectrum direction θ', and the *y*-axes correspond to the direction of observed spectrum, θ.

The fitted parameters are plotted in Figure 33a where θ_p varies between -5° and 5°, and the mean direction over all depths is close to zero. Values of σ_θ vary between 8° and 20° and is a function of depth, which confirms that the covariance density in shallow water is concentrated in few directions. The direction $\theta = 0°$ corresponds to the direction normal to the boundary.
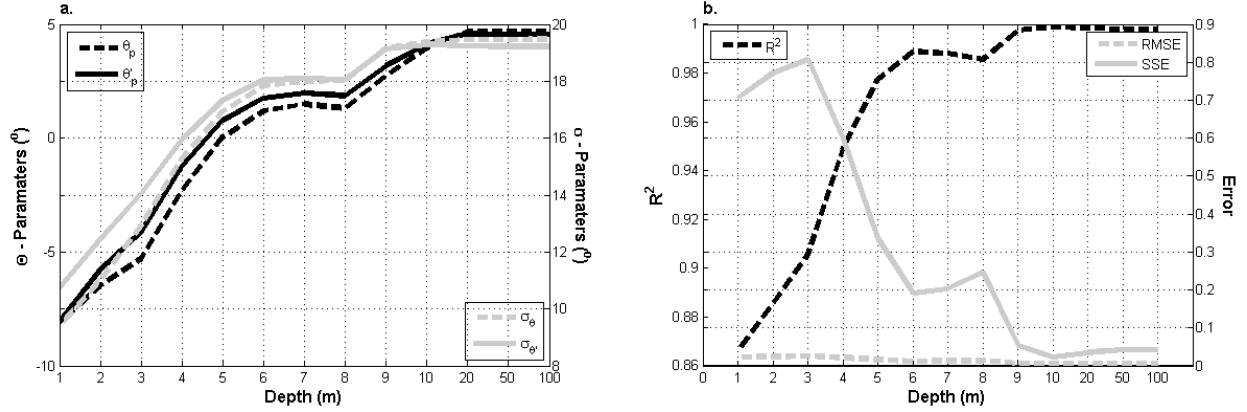
Figure 33. a. Parameters of Eq. (3.25) for each depth, and b. goodness-of-fit statistics of the function (3.25) for each depth.

The goodness-of-fit statistics of the covariance function in direction are presented in Figure 33b. From the plot, it is evident that for all the depths the suggested covariance function accurately models the covariance for each depth. The correlation and the error are reduced with the increase in depth. The relatively limited accuracy of the covariance function for $d \leq 4m$ is due to the asymmetric directionality.
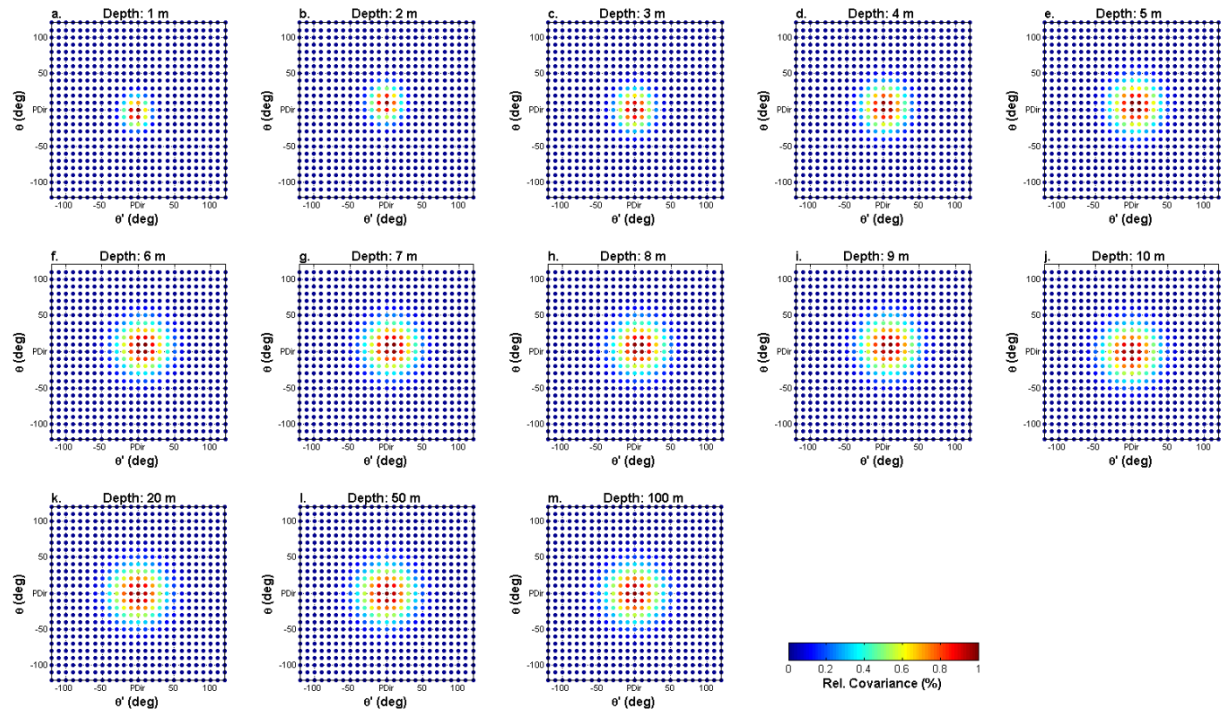
Figure 34. Estimated normalized covariance in direction, $C_{\theta\theta'}$, based on the function (3.25) for thirteen different depths; the x-axes correspond to θ', and the y-axes correspond to the θ.

The covariance function in direction is examined in comparison with the calculated covariance. In Figure 34, the estimated covariances are plotted. Figure 32 and Figure 34, show that the covariance density from the covariance function is more spread in direction than the calculated

one for $d \leq 4m$. The covariances for depths $d \geq 5m$ have similar directional distribution, and the differences for depths $d \geq 8m$ are eliminated.

y = 1.037x -0.002
R^2 = 0.983
RMSE = 0.015
SSE = 3.766

With y-axis labeled "Covariance Function" and x-axis labeled "Normalized Covariance".
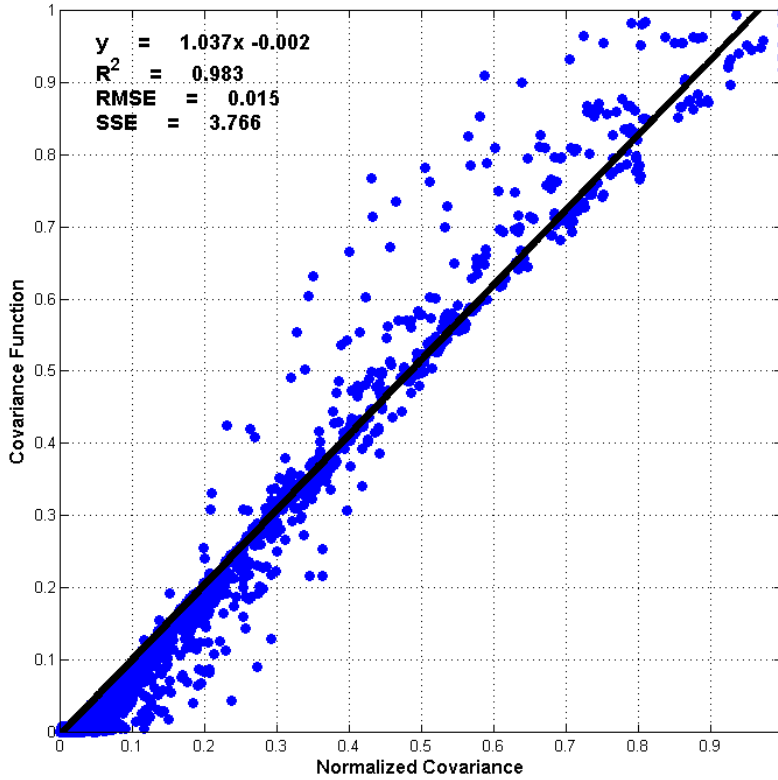
Figure 35. Scatter plot of the calculated covariance for all the depths versus the estimated values of the covariance function.

These observations are summarized in the scatter plot of the covariance for all the depths between the calculated and the estimated from Eq. (3.25). At high values of normalized covariance, $C_{\theta\theta'} > 0.3$, the covariance function overestimates in few cases by up to 25%, due to the broadness of the covariance function. At low covariance values, $C_{\theta\theta'} < 0.1$, the covariance function underestimates the covariance. More precisely, for values of the covariance close to zero, the covariance function predicts smaller covariance values. By considering the general properties of Eq. (3.25), it is clear the correlation length in direction, σ_θ and $\sigma_{\theta'}$, could be slightly smaller for $d \leq 4m$ and larger for $d > 4m$, as evidenced Figure 36. In general, however, the covariance function accurately models the covariance, and the values of the error statistics are small. The regression function between the two covariances show lack of systematic bias.
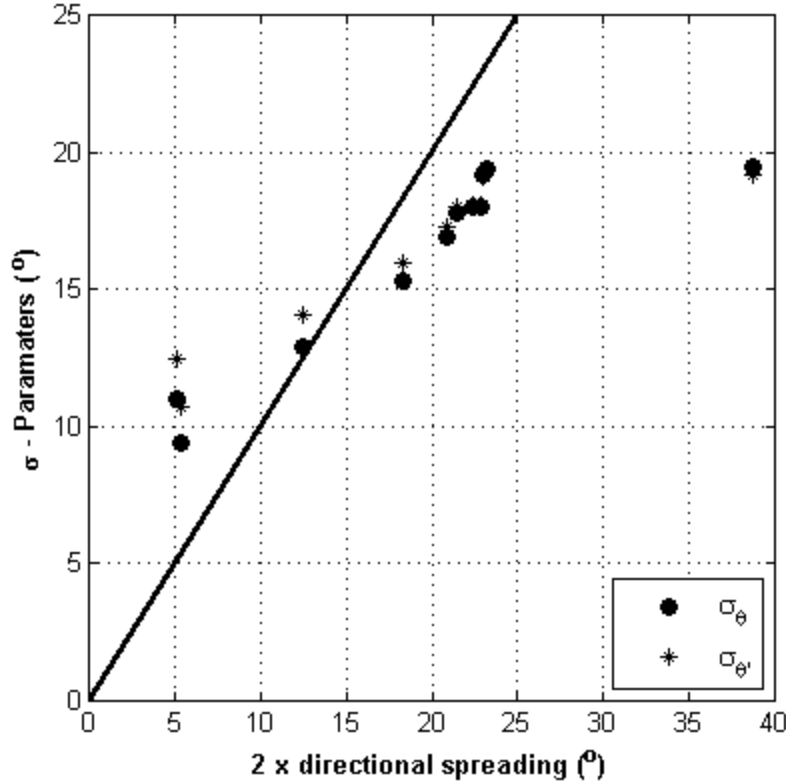
Figure 36. Scatter plot of the averaged directional spreading of the covariance spectra in comparison the σ_θ parameters of equation (3.25). The solid line is $y = x$.

3.3.2.3 Error Covariance Function $C_{f'\theta'}$

Throughout the development of the covariance function, it is assumed that the covariance is independent in the different dimensions of the model. From Figure 20 and Figure 21, it is evident that at least at the spectral space, the covariance is a function of frequency and direction. Therefore, a covariance function in frequency f' and direction θ' has to be addressed as well. The covariance spectra follow positively skewed normal distribution in frequency and near normal distribution in direction. Therefore, $C_{f'\theta'}$ is assumed to be a skew normal distribution (Azzalini and Capitanio, 2014):

$$C_{f'\theta'} = \exp\left(-\left(\left(\frac{f' - \alpha_{f'}}{2b_{f'}}\right)^2 + \left(\frac{\theta' - \alpha_{\theta'}}{2b_{\theta'}}\right)^2\right)\right) \text{erf}(\lambda_{f'}f' + \lambda_{\theta'}\theta') \qquad (3.26)$$

where $a_{f'}, b_{f'}, \lambda_{f'}, a_{\theta'}, b_{\theta'}, \lambda_{\theta'}$ are parameters to be fitted in frequency and direction. In this case, the apparent advantage of using Eq. (3.26) is the modeling of the skewness in the frequency domain and to have the advantage of the physical interpretation of each of the factors of the function. The input data have been integrated over f and they depend on θ, θ' and f'. Following the same basic steps as in the previous sections, Eq. (3.26) is fitted to the normalized covariance spectra for all the available θ. Hence, for each direction there is a set of 6 parameters. The

49

goodness-of-fit statistics for all the fits as a function of the direction and depth are illustrated in Figure 37. Equation (3.26) was fitted for data of directions between $-100° \leq \theta \leq 100°$ and all the depths. The data and the fit do not have significant correlation for the directions where there is not significant covariance density (Figure 31) due to the angle between the wave field propagation direction and the directions under examination $> 70°$. In addition, the reduced accuracy of the covariance for shallow grids is the result of the combination of two phenomena: the reduction of the directional spreading and the concentration of the wave energy density, and consequently of the covariance density, at few frequencies (as seen in e.g. Figure 32a). Therefore, the fitted data are filtered based on Table 4 and R^2, shown in Figure 37a.
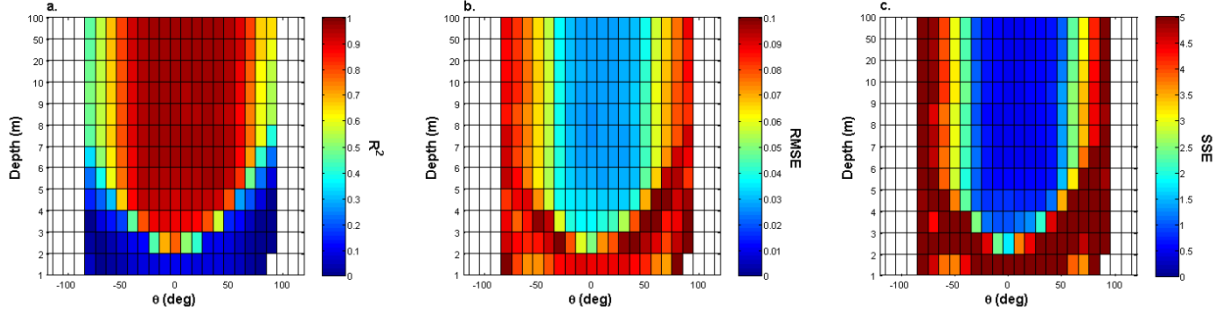

Figure 37. a. R^2, b. $RMSE$ and c. SSE as a function of the direction, θ, and depth for the fit of Eq. (3.26) to all the available covariance spectra.

The parameters in Eq. (3.26) are visualized in Figure 38 as a function of direction θ and depth. All the six parameters depend on the direction θ, and their values are almost constant for $\theta = 0°$ (normal to the boundary).They vary as a function of depth. Specifically, $a_{f'}$ (Figure 38a) varies between 0.09 and 0.11, with mean value 0.094. $b_{f'}$(Figure 38b) varies between 0.01 and 0.024 with mean value 0.176. Considering that $b_{f'}$ expresses standard deviation in frequency, the observed high values for large angles are expected, but for $\theta = 0°$, $b_{f'}$ is almost constant with mean value 0.144. $l_{f'}$ (Figure 38c) varies between 0 and 1.2. The small values of $\lambda_{f'}$ impose zero skewness in frequency and consequently, the covariance in frequency is symmetric. This behavior strongly depends on the direction θ and $l_{f'} \approx 0$ for $\theta = |35°|$. The $a_{\theta'}$ and $b_{\theta'}$ (Figure 38d and e) confirm the previous findings that $a_{\theta'} \leq 0.5d\theta$ and the $b_{\theta'} \leq 0.5D_{sp}$. Finally, the $\lambda_{\theta'}$(Figure 38f) is three to four orders of magnitude smaller than the actual values of the variable and is, therefore, set to zero.

Summarizing these conclusions, for $\theta = |d\theta/2|$, Eq. (3.26) becomes:

$$C_{f'\theta'} = \exp\left(-\left(\left(\frac{f' - 0.091}{2*0.13}\right)^2 + \left(\frac{\theta' - 0.5d\theta}{2*20}\right)^2\right)\right)\mathrm{erf}(\lambda_{f'}f' + 10^{-3}\theta')$$

(3.27)

$$\lambda_{f'} = \begin{cases} -0.163d + 2.059, & d \leq 10m \\ 4.29, & d > 10m \end{cases}$$
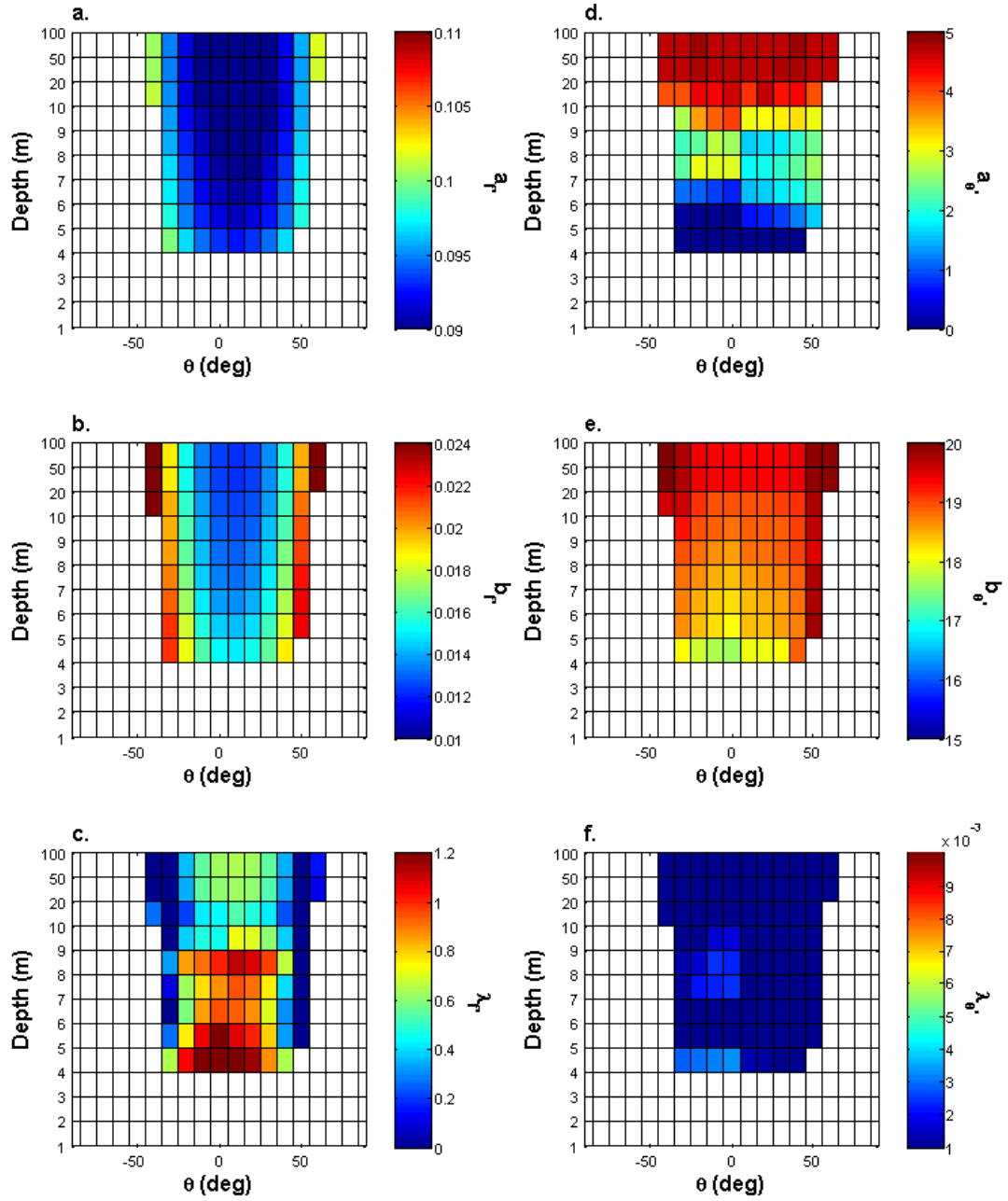
Figure 38. The values of the fitted parameters a. $a_{f'}$, b. $b_{f'}$, c. $\lambda_{f'}$, d. $a_{\theta'}$, e. $b_{\theta'}$, and f. $\lambda_{\theta'}$ of Eq. (3.26).

For $\theta = 0°$ the scatter plot of the covariance function and its statistical measures of fit are given in Figure 39.
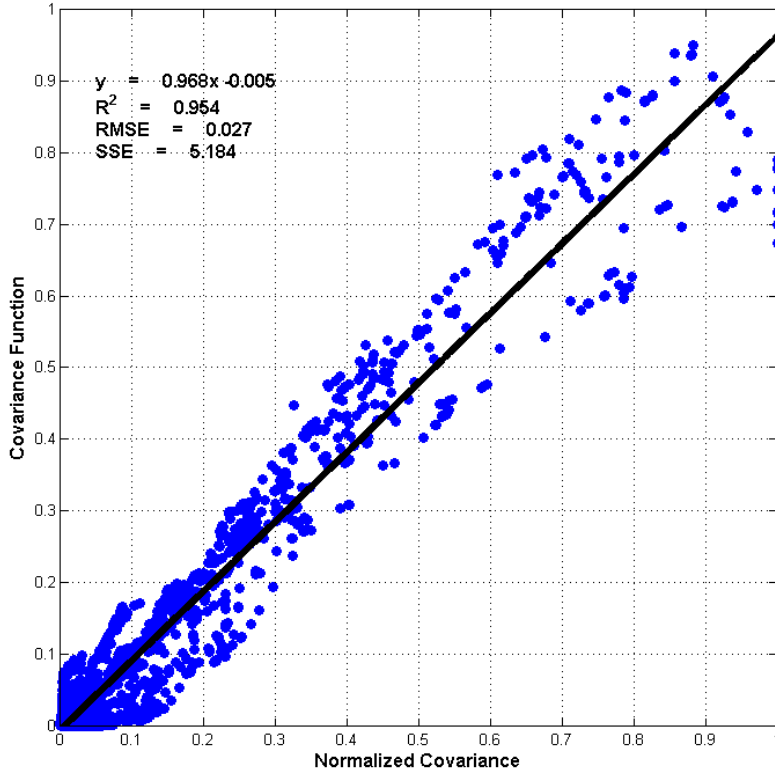
Figure 39. Scatter plot of the calculated covariance for all the depths versus the estimated values of the covariance function $C_{f\theta'}$.

3.4 Effect of Wind on the Covariance in Space as a Function of Frequency

As described in the introduction to section 3.3, in SWAN the wind energy is transferred to the waves near the peak of the spectrum and at the mid-range frequencies. In order to determine the effect of the wind on the covariance, the covariance matrices from the $10m$ homogenous grid, forced with only the wave field are compared with covariance of the ensemble forced with the waves and the wind. The wind data used in this case is the measured wind during the wave measurements used as boundary conditions as shown in Figure 3.

The effect of the wind is determined by comparing the maximum covariance of the two ensembles for each frequency along the wave propagation direction (Figure 40). For frequencies up to 0.0612 Hz, the two covariances are almost identical over space. At frequency, $f = 0.0691$ Hz, a second peak can be seen, which becomes more evident between the frequencies $0.0780\ Hz \leq f \leq 0.1427\ Hz$. The second peak vanishes between the frequencies $0.161 Hz \leq f \leq 0.2050\ Hz$. For frequencies higher than 0.23 Hz, wind seems to have a significant effect. However, the magnitudes of the covariance for these frequencies are up to four orders smaller than the covariance at lower frequencies. Thus, it can be concluded that there is no significant wind effect on the covariance for nearshore domains of a few kilometers and can be neglected.
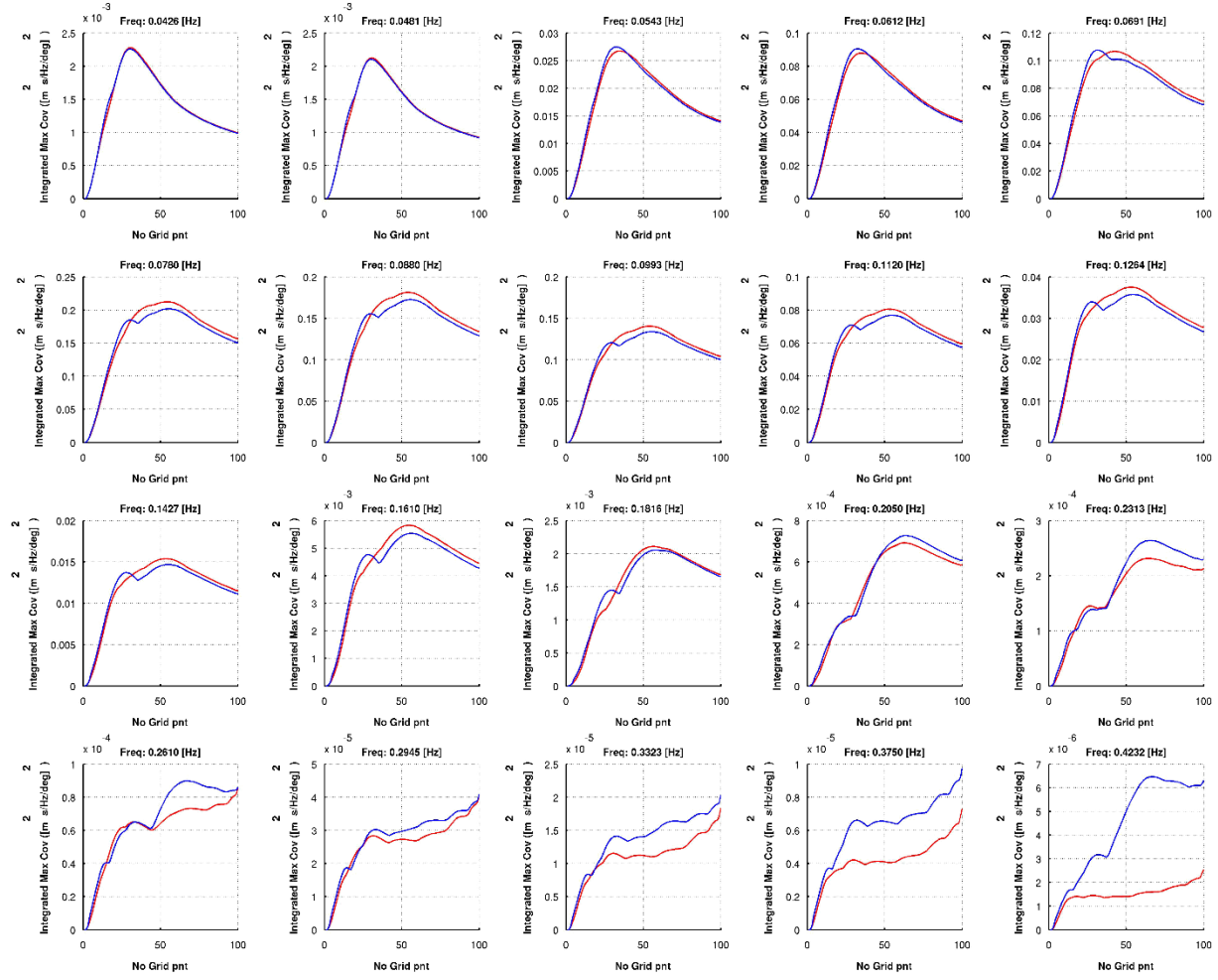
Figure 40. Spatial distribution of the maximum covariance density for each frequency. The one ensemble is forced a. only with waves (red) and b. with waves and wind (blue).

4   CONCLUSIONS

This study develops a covariance function for nearshore wave data assimilation systems driven by the data and the wave model SWAN. Under the basic assumption that the best estimator of model covariance error is the mean of an ensemble of forward simulations or measurements, the covariance matrices in all the dimensions of the problem were calculated and then were modeled. For the development of the covariance function, each dimension of the problem was modeled independently, with the one exception where there was correlation between the wave direction and frequency.

It was found that the temporal covariance can be modeled by a parameterized Gaussian function. The temporal correlation length of the function depends on the local conditions and the temporal length of the modeled phenomena, climatological, seasonal and extreme events, such as storms. The temporal length for long period studies is on the order of days, but for the extreme events is almost constant with average value 6 hours. The temporal length has been determined in the two

temporal scales, for all the locations where NDBC has wave measuring sensors. This information can be summarized in a look-up table, available to be integrated in any wave assimilation system. In addition, the function was optimized by examining the properties of the parameters of the temporal covariance function.

For nearshore wave assimilation applications, the covariance function depends primarily on the local depth and secondly on the distance from the assimilation location. The basic concept of the covariance function is the binary categorization of the wave field in breaking or non-breaking. For non-breaking wave field, the error covariance is approximately constant as a function of the distance, but it strongly depends on the local depth with the correlation length in distance inversely proportional of the bottom slope. For breaking waves, the covariance values decrease exponentially in very short distance.

The challenging part of this study is the determination of the covariance in the spectral space, and more specifically in frequency, due to the dominant non-linear processes or even the numerical implementation in SWAN of the wave related processes. A conceptual model of the behavior of the covariance in frequency of a shoaling wave field was developed. Due to the data volume, the covariance are binary separated into significant and non-significant, according to the covariance density. It was determined that the covariance spectrum in frequency is positively skewed. Therefore, the suggested covariance functions can model the observed long tail function of frequency with a sum of normal distribution and skewed normal distribution. The covariance in direction is symmetric; hence, a two dimensional exponential function is validated and suggested for covariance function in direction. By combining the properties of the covariance in spectrum, a skewed exponential function, similar to the normal distribution is suggested.

In short, the simplest suggested error covariance function for a 4D-var wave assimilation system for wave spectra is the following:

$$C_f(x, y, \theta, f, t, x', y', \theta', f', t') = V C_{xx'} C_{yy'} C_{f'\theta'} C_{tt'}$$

$$C_f = V\sqrt{2\pi b_t} \exp\left(-\left[\frac{(t-a_t)}{b_t}\right]^2\right) \left(\exp\left(-\frac{|x-x'|^{\beta_x}}{L_x}\right) + \exp\left(-\frac{|y-y'|^{\beta_y}}{L_y}\right)\right)$$

$$10^{\left(-1.54\exp\left(-\frac{d+1.25}{5.33}\right)^2\right)} \exp\left(-\left(\left(\frac{f'-0.091}{2*0.13}\right)^2 + \left(\frac{\theta'-0.5d\theta}{2*20}\right)^2\right)\right) \mathrm{erf}\left(\lambda_{f'} f'\right.$$

$$\left. + 10^{-3}\theta'\right)$$

(4.1)

5   OUTLOOK

The determination of covariance in assimilation systems is an open question, which has important impact on the data assimilation systems. In this investigation, an optimized covariance

function for all the wave model dimensions is introduced. The function is based on the analysis of simple forward simulations and on some of the available wave measurements around US coasts. The assumptions of this approach has apparent disadvantages, but the resulting function is the basis for more accurate assimilation systems. The near-future research in the subject is focused on the extended validation, optimization and generalization of the error covariance function. In order to increase the accuracy and efficiency of the covariance function, the next step should be the systematic and operational assessment of the statistical properties of the wave field simulation errors on a global scale, by integrating in the analysis of all available data with their error, or inversely, their accuracy. The basic step to understand and model the error of the wave models is to combine the existing simulations with the measurements.

From the simulations side, there are operational wave simulations in global, regional and local scale. The two most broadly used are the simulations from NOAA, based on WW3 and from ECMWF based on WAM. In addition to these, the most coastal countries provide operationally local wave simulations tailored to the local environment and conditions, based mainly on SWAN and WW3.

From the measurements side, there are available in situ data in global scale, for instance the member countries of European Union have started creating common databases with available data from the Norwegian Sea to the East Atlantic and most of the N. Mediterranean Sea. In addition, after the Deepwater Horizon accident, all the oil companies with offshore operations collaborate with governmental authorities and international organizations and collect basic oceanographic data. For instance, the SIMORC system has in situ oceanographic data from remote areas, such as West Central Africa. Moreover, the expansion of ground based remote sensing, mainly the High Frequency radars, offers unique data sets of time series of wave parameters with spatial coverage of few hundred meters. In addition, the satellite remote sensing offers stable wave products from SAR and Altimeter sensors. An open subject is the combination of all the available datasets and common standardization with defined accuracy for each of them.

It is clear that the information from the measurements and the wave models is available, but to further develop the error covariance function, it is necessary to process the information in a consistent and systematic way. Such a process will need to be able to ingest large, multidimensional wave datasets and have the ability to automatically update the error covariance function using advanced techniques of machine learning and data mining.

6   REFERENCES

Almeida, T.G., Walker, D. and Reath, A.L., 2014. Estimation of waves, tides and currents near the new river inlet using variational inverse modeling. Ocean Sciences Meeting, Honolulu, Hawaii USA.

Aouf, L., Lefèvre, J.-M. and Hauser, D., 2006. Assimilation of directional wave spectra in the wave model wam: An impact study from synthetic observations in preparation for the swimsat satellite mission. Journal of Atmospheric and Oceanic Technology, 23(3): 448-463.

Azzalini, A. and Capitanio, A., 2014. The skew-normal and related families, 1. Cambridge University Press.

Bennett, A.F., 2002. Inverse modeling of the ocean and atmosphere. Cambridge University Press, Cambridge, UK ; New York, xxii, 234 p. pp.

Bennett, A.F., Chua, B.S. and Leslie, L.M., 1996. Generalized inversion of a global numerical weather prediction model. Meteorol Atmos Phys, 60(1-3): 165-178.

Berre, L. and Desroziers, G., 2010. Filtering of background error variances and correlations by local spatial averaging: A review. Monthly Weather Review, 138(10): 3693-3720.

Booij, N., Ris, R.C. and Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions - 1. Model description and validation. J Geophys Res-Oceans, 104(C4): 7649-7666.

Box, G.E.P. and Jenkins, G.M., 1970. Time series analysis; forecasting and control. Holden-day series in time series analysis. Holden-Day, San Francisco,, xix, 553 p. pp.

Breivik, L.A. and Reistad, M., 1994. Assimilation of ers-1 altimeter wave heights in an operational numerical wave model. Weather Forecast, 9(3): 440-451.

Breivik, L.A. et al., 1998. Assimilation of ers sar wave spectra in an operational wave model. J Geophys Res-Oceans, 103(C4): 7887-7900.

Chen, H.S., Behringer, D., Burroughs, L.D. and Tolman, H.L., 2004. A variational wave height data assimilation system for ncep operational wave models NWS/NOAA, Camp Springs, Maryland.

Evensen, G., 2009. Data assimilation : The ensemble kalman filter. Springer, Dordrecht ; New York, xxiii, 307 p. pp.

Feng, X.B., Zheng, J.H. and Yan, Y.X., 2012. Wave spectra assimilation in typhoon wave modeling for the east china sea. Coastal Engineering, 69: 29-41.

Fisher, M., 2003. Background error covariance modelling. ECMWF Seminar on Recent Developments in Data Assimilation, Reading, United Kingdom, p.^pp. 45-63.

Flampouris, S., 2003. Improvement of wave field predicition of wam, University of the Aegean, Mytilene, Greece, 96 pp.

Flampouris, S., Veeramony, J., Orzech, M. and Ngodock, H., 2014. Development of the error covariance function for assimilation of wave spectra – the example of swan, Ocean Sciences Meeting, Honolulu, Hawaii USA.

Flampouris, S., Veeramony, J., Orzech, M.D. and Ngodock, H.E., 2013. Validation of a wave data assimilation system based on swan. European Geosciences Union General Assembly, Vienna, Austria, p.^pp.

Greenslade, D.J.M., 2001. The assimilation of ers-2 significant wave height data in the australian region. J Marine Syst, 28(1-2): 141-160.

Hasselmann, K., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). Ergänzungsheft zur deutschen hydrographischen zeitschrift reihe a (8^0), nr 12. Deutsches Hydrographisches Institut, Hamburg,, 95 p. pp.

Holthuijsen, L.H., 2007. Waves in oceanic and coastal waters. Cambridge University Press, Cambridge, xvi, 387 p. pp.

Komar, P.D., 1998. Beach processes and sedimentation. Prentice Hall, Upper Saddle River, N.J., x, 544 p. pp.

Komen, G.J. et al., 1994. Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge ; New York, NY, USA, xxi, 532 p. pp.

Kuik, A.J., Vanvledder, G.P. and Holthuijsen, L.H., 1988. A method for the routine analysis of pitch-and-roll buoy wave data. Journal of Physical Oceanography, 18(7): 1020-1034.

Lionello, P., Guenther, H., Hansen, B., Janssen, P.A.E.M. and Hasselmann, S., 1994. An optimal interpolation scheme for assimilating altimeter data into wan model. In: G.J. Komen et al. (Editors), Dynamics and modelling of ocean waves. Cambridge University Press, Cambridge ; New York, NY, USA.

Menke, W., 2012. Geophysical data analysis : Discrete inverse theory. Academic Press, Waltham, MA, 293 p. pp.

Ngodock, H. and Carrier, M., 2014. A 4dvar system for the navy coastal ocean model. Part i: System description and assimilation of synthetic observations in monterey bay. Mon. Wea. Rev, 142: 2085-2107.

Ngodock, H.E., 2005. Efficient implementation of covariance multiplication for data assimilation with the representer method. Ocean Modelling, 8(3): 237-251.

Nichols, N.K., 2010. Mathematical concepts of data assimilation. In: W. Lahoz, B. Khattatov and R. Menard (Editors), Data assimilation. Springer Berlin Heidelberg, pp. 13-39.

Orzech, M., Veeramony, J. and Flampouris, S., 2014. Swanfar: A 4dvar data assimilation system for swan. Ocean Sciences Meeting, Honolulu, Hawaii USA, p.^pp.

Orzech, M.D., Veeramony, J. and Ngodock, H., 2013. A variational assimilation system for nearshore wave modeling. Journal of Atmospheric and Oceanic Technology, 30(5): 953-970.

Parvaresh, A., Hassanzadeh, S. and Bordbar, M.H., 2005. Statistical analysis of wave parameters in the north coast of the persian gulf. Ann Geophys-Germany, 23(6): 2031-2038.

Rasmussen, C.E. and Williams, C.K.I., 2006. Gaussian processes for machine learning. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., xviii, 248 p. pp.

Ris, R.C., Holthuijsen, L.H. and Booij, N., 1999. A third-generation wave model for coastal regions - 2. Verification. J Geophys Res-Oceans, 104(C4): 7667-7681.

Sannasiraj, S.A., Babovic, V. and Chan, E.S., 2006. Wave data assimilation using ensemble error covariances for operational wave forecast. Ocean Modelling, 14(1-2): 102-121.

Soares, C.G., Ferreira, A.M. and Cunha, C., 1996. Linear models of the time series of significant wave height on the southwest coast of portugal. Coastal Engineering, 29(1-2): 149-167.

Spanos, P.T.D., 1983. Arma algorithms for ocean wave modeling. J Energ Resour-Asme, 105(3): 300-309.

Thornhill, G.D. et al., 2012. Integration of a 3d variational data assimilation scheme with a coastal area morphodynamic model of morecambe bay. Coastal Engineering, 69(0): 82-96.

Tolman, H.L., 2002. Alleviating the garden sprinkler effect in wind wave models. Ocean Modelling, 4(3-4): 269-289.

Tolman, H.L., 2009. User manual and system documentation of Wavewatch III version 3.14. 276, NOAA/NWS/NCEP/MMAB, Camp Springs, Maryland.

Torsethaugen, K., 1994. Model for a doubly peaked spectrum. Lifetime and fatigue strength estimation implications. In: E. Watanabe and E. Bratteland (Editors), International Workshop

on Floating Structures in Coastal zone. Port and Harbor Research Institute, Yokosuka, Japan, Hiroshima, pp. 574.

Veeramony, J., Walker, D. and Hsu, L., 2010. A variational data assimilation system for nearshore applications of swan. Ocean Modelling, 35(3): 206-214.

Voorrips, A.C., Makin, V.K. and Hasselmann, S., 1997. Assimilation of wave spectra from pitch-and-roll buoys in a north sea wave model. J Geophys Res-Oceans, 102(C3): 5829-5849.

Walker, D.T., 2006. Assimilation of sar imagery in a nearshore spectral wave model.

Wunsch, C., 2006. Discrete inverse and state estimation problems : With geophysical fluid applications. Cambridge University Press, Cambridge ; New York, xi, 371 p., 12 p. of plates pp.

Yim, J.Z., Chou, C. and Ho, P., 2002. A study on simulating the time series of siginificant wave height near the keelung harbor. International Society of Offshore and Polar Engineers Conference, p.^pp. 92-96.

Zupanski, D., 1997. A general weak constraint applicable to operational 4dvar data assimilation systems. Monthly Weather Review, 125(9): 2274-2292.